2000

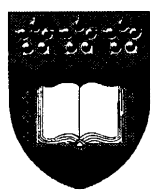
Studies in reverse transcriptase: synthesis of inhibitors, conformational analysis, and integration of drug design

Tien T.T. Luu
University of Wollongong

Studies of Reverse Transcriptase : Synthesis of Inhibitors, Conformational Analysis, and Integration of Drug Design

Tien T. T. Luu, B.Med.Chem (Hons.)

A thesis submitted in fulfilment of the
requirements for the degree of Doctor of Philosophy



Department of Chemistry

University of Wollongong
Wollongong, AUSTRALIA

December, 2000

For Mum and Dad

Acknowledgements

To my supervisors: Paul, who provided the means for me to do more than just join the circus, and in his own way has probably suffered as much as I have; and Renate, for all her care and enduring support even when I was in the wrong, she has been an inspirational role model and I am forever indebted.

To my parents, whose love and support has always been, and will always be, endless and immeasurable. I could never have come so far without them. To my brothers, Trung and Tuan, and to Hoa for being there all the time, and resuscitating all the computers I somehow managed to destroy.

To Michelle, my constant anchor in stormy seas, who has guided me, believed in me, suffered with me, and I will never be the same again.

To Glenn, with whom I have shared this rite of passage; the journey would have been so much more bleak were it not for his support and friendship.

To Dan the Man, Helen, Jody and all past and present members of the Keller research group for the great friendships, and stimulating conversations that I will truly miss.

To Stephen Titmuss and Scott Leach for their great work with the difference distance matrices upon which the conformational studies were based.

To Dr Sebastian Marcuccio and Dr Mary Rodopoulos from CSIRO Molecular Science for making available to us the bis(pinacolato)diboron necessary for the Suzuki reactions, and for their helpful insight into this reaction.

To Tripos Inc. and Dr Chris van Dyke for the trial of the Tripos software and constructive help for all our questions; without this generosity our work with the superligand would have been much harder to realise.

To AMRAD Operations and Professor Yodhathai Thebtaranonth for conducting the antiviral and antimalarial testing, respectively.

Abstract

The reverse transcriptase (RT) enzyme of the human immunodeficiency virus is an attractive target for rational drug design as it is an essential and unique enzyme in the viral life cycle. RT can be inhibited by non-nucleoside inhibitors acting at an allosteric binding pocket (NNIBP). The inherent flexibility of the enzyme has made the study of the mechanism(s) of its functions and inhibition by non-nucleoside inhibitors difficult and ambiguous. Furthermore, this flexibility in the NNIBP has made it difficult to characterise the binding pocket for rational drug design.

The synthesis of simple analogues of the antiviral michellamine alkaloids was undertaken to determine if they were active at the NNIBP, as the michellamines are known to inhibit RT. A Suzuki cross-coupling methodology was found to be the most efficient and high yielding route to substitute simple aromatic aryl and alkyl moieties in place of the tetrahydroisoquinoline moiety of michellamine. It was found that a weak base was required to form the boronate ester, but a stronger base was necessary for the subsequent cross coupling to proceed. The homo-coupling to synthesise the dimeric series of analogues was performed under essentially the same conditions, by forming the boronate ester of the *ortho*-brominated monomeric analogue and self-coupling to achieve dimerisation, in moderate yield. Deprotection of the synthesised monomeric and dimeric analogues was performed to yield the desired potential inhibitors. These were sent for biological testing, with results unavailable at the time of printing.

An objective, and quantitative method to analyse conformational changes of the reverse transcriptase enzyme induced by substrate and inhibitor binding was established. Using this technique, a large number of substrate and non-nucleoside inhibitor-bound x-ray crystallographic structures were superimposed onto an unliganded form of RT to study the conformational displacement and change of the polymerase and RNase H active sites and the primer and template grips, upon substrate and inhibitor binding. The binding of inhibitors to the NNIBP, in particular, induced large conformational changes, both proximal to and remote from the binding pocket. The residues of the polymerase and RNase H active sites showed small conformational displacements in substrate

structures, possibly to adopt catalytically active conformations to recognise the nucleic acid and incoming nucleotide, and prepare for polymerisation and ribonuclease hydrolysis. The template and primer grips in the substrate bound structures also showed displacements to maximise intermolecular interactions with the bound double stranded DNA. In the inhibitor-bound structures the active sites showed large displacements that affected the possible coordination complexes of the catalytic residues with the template and primer strands, the incoming nucleotides, and essential magnesium cations. The nucleic acid grips were displaced into a space occupied by the double stranded DNA in the substrate-bound structures, and would either prevent nucleic acid from binding to RT, or disrupt the normal intermolecular interactions between the enzyme and the nucleic acid. The results supported literature postulating that a mechanism of inhibition involved an opening of the hand-shaped structure of the enzyme as the structural motifs and active sites moved away from each other and the bound nucleic acid.

Computer aided ligand design has traditionally been segregated into the classification of either ligand-based or structure-based. Preliminary studies were undertaken to establish an integrated approach to drug design using computational tools from both strategies. Ligands and structures for both approaches were deliberately limited to those from x-ray crystallographic structures. Ligand-derived pharmacophores were generated, merged and analysed to determine ability to correlate ligand structure to activity and predictive power. A structure-derived pharmacophoric query was developed from the activity 'weighted' model of the three-dimensional x-ray structures whose inhibitors were used in the ligand-based pharmacophore. This query, termed the 'superligand', represented all the potential binding interactions of residues of the active site to a potential ligand in the pocket. Database searching with this superligand identified several compounds of interest, which were obtained for biological testing. The superligand was converted to a traditional pharmacophore form and represented the first step toward the generation of the 'combiphore', an integration of ligand- and structure-based pharmacophores.

Abbreviations

Amino acid residues

amino acid	three letter code	single letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

(-)ssDNA	minus strand strong stop DNA
(+)ssDNA	plus strand strong stop DNA
Ar	aryl
$Ba(OH)_2$	barium hydroxide
bd	broad doublet
CA	capsid protein
$CDCl_3$	deteurated chloroform
CI	chemical ionization (mass spectrometry)
cPPT	central polypurine tract
CTS	central termination sequence
d	chemical shift in parts per million downfield from tetramethylsilane
d	doublet (spectral)
D+	downstream plus strand DNA
DCM	dichloromethane
DME	ethylene glycol dimethyl ether
DMF	dimethylformamide
DNA	deoxyribonucleic acid

dNTP	deoxynucleotide triphosphate
DTP	Development Therapeutics Program (NCI)
EI	electron impact (mass spectrometry)
Et	ethyl
EtOH	ethanol
g	gram(s)
HAART	highly active anti-retroviral therapy
HIV	human immunodeficiency virus
HMBC	heteronuclear multiple bond coherence
hr	hour(s)
HRMS	high resolution mass spectrometry
HSQC	heteronuclear single quantum coherence
Hz	hertz
IN	integrase enzyme
J	coupling constant (NMR spectroscopy)
K_2CO_3	potassium carbonate
KOAc	potassium acetate
L	litre(s)
LTR	long terminal repeat
μ	micro
m	multiplet (NMR spectroscopy), milli
M	moles per litre
mp	melting point °C
m/z	mass to charge ratio (mass spectrometry)
MA	matrix protein, microanalysis
Me	methyl
MeOH	methanol
MHz	megahertz
min	minute(s)
mol	mole(s)
mRNA	messenger RNA
MS	mass spectrometry
NC	nucleocapsid protein
NCI	National Cancer Institute
NMR	nuclear magnetic resonance
NNIBP	non-nucleoside inhibitor binding pocket
NOESY	nuclear Overhauser effect spectroscopy
°C	degrees Celsius
p51	51 kDa subunit of HIV-1 reverse transcriptase
p66	66 kDa subunit of HIV-1 reverse transcriptase
PAS	polymerase active site
PBS	primer binding site
$Pd(PPH_3)_4$	tetrakis(triphenylphosphine) palladium(0)
PDB	Protein Data Bank
$PdCl_2$(dppf)	dichloro[1,1'-bis(diphenylphosphino)ferrocene]palladium
PP_i	pyrophosphate
ppm	parts per million (NMR spectroscopy)
PPT	polypurine tract
PR	protease enzyme
q	quartet (NMR spectroscopy)

R	repeat
Rf	retention factor (chromatography)
RMS	root mean square
RNA	ribonucleic acid
RNase H	ribonuclease hybrid
RCSB	Research Collaboratory for Structural Bioinformatics
RT	reverse transcriptase
s	singlet (NMR spectroscopy)
SAR	structure-activity relationship
SU	surface glycoprotein
t	triplet (NMR spectroscopy)
THF	tetrahydrofuran
TLC	thin layer chromatography
TM	transmembrane protein
tRNA	transfer RNA
U+	upstream plus strand DNA

Table of Contents

Declaration iii
Acknowledgements iv
Abstract v
Abbreviations vii
Table of Contents x

Chapter 1 - Introduction

1.1 The Human Immunodeficiency Virus 1
1.1.1 HIV-1 genome and virion structure 2
1.1.2 HIV life cycle 4
1.1.3 Immune response to HIV 5
1.1.4 Virus reservoirs and the mutability of HIV 6
1.1.5 Nucleoside reverse transcriptase inhibitors (NRTIs) 8
1.1.6 Non-nucleoside reverse transcriptase inhibitors (NNRTIs) 11
1.1.7 Natural products as antivirals 14
1.2 The Michellamines 17
1.2.1 Structure of the Michellamines and the Korupensamines 18
1.2.2 Synthesis of the michellamines 19
1.2.3 Analogues of the michellamines; antiviral and antimalarial activity 23
1.3 Computational chemistry of biological and chemical systems 25
1.4 Research aims 29
1.4.1 Synthetic design objectives 29
1.4.2 Molecular modelling objectives 30
1.4.3 Drug design objectives 31

Chapter 2 - Synthesis of Michellamine Analogues

2.1 Introduction 32
2.1.1 Inhibitor design and strategy 32
2.2 Biaryl bond formation 35
2.2.1 Stille reaction 36
2.2.2 Suzuki reaction 38
2.3 Synthesis of the brominated naphthyl unit (65) 41
2.4 Cross-coupling 42
2.4.1 Stille cross-coupling reactions 43
2.4.1.1 Preparation of stannanes 44
2.4.1.2 Coupling of naphthyl core halide and aryl stannanes 47
2.4.2 Suzuki cross-coupling reactions 48
2.4.2.1 Preparation of boronic acids and esters 49
2.4.2.2 Coupling of naphthyl halides and aryl boronic acids 52
2.4.2.3 Coupling of naphthyl boronate esters and aryl halides 57
2.5 Ortho-directed bromination 63
2.6 Dimerisation reactions 65
2.6.1 Stille homo-couplings 65
2.6.2 Suzuki homo-couplings 66
2.7 Deprotection reactions 71

2.8 Transformations of the ethyl ester 73
2.9 Antiviral and antimalarial testing 76
2.10 Future synthetic targets 76

Chapter 3 - Conformational Studies of Reverse Transcriptase

3.1 Reverse transcriptase 78
3.1.1 Structure of HIV-1 reverse transcriptase 78
3.1.2 Reverse transcription 80
3.1.3 The non-nucleoside inhibitor binding pocket (NNIBP) and inhibition of RT 84
3.2 Superimposition 88
3.2.1 Difference distance matrices 88
3.2.2 Derivation of the superimposition subset 90
3.2.3 Comparison of superimposition techniques 94
3.2.4 Superimposition for conformational studies 97
3.3 The polymerase active site 98
3.3.1 The polymerase active site and polymerisation 98
3.3.2 Synopsis of the displacement of the PAS 101
3.3.2.1 Substrate induced conformational changes of the PAS 103
3.3.2.2 Inhibitor induced conformational changes of the PAS 106
3.3.3 Conformational changes for aspartic acid 110 107
3.3.4 Conformational changes for tyrosine 183 109
3.3.5 Conformational changes for methionine 184 111
3.3.6 Conformational changes for aspartic acid 185 114
3.3.7 Conformational changes for aspartic acid 186 116
3.3.8 Activation and inhibition of the PAS 117
3.4 The primer grip 120
3.4.1 The primer grip and its role in reverse transcription 120
3.4.2 Synopsis of the displacement of the primer grip 122
3.4.2.1 Substrate induced conformational changes of the primer grip 124
3.4.2.2 Inhibitor induced conformational changes of the primer grip 126
3.4.3 Conformational changes for phenylalanine 227 128
3.4.4 Conformational changes for tryptophan 229 130
3.4.5 Conformational changes for methionine 230 131
3.4.6 Activation and inhibition of the primer grip 133
3.5 The template grip 138
3.5.1 The template grip and its role in reverse transcription 138
3.5.2 Synopsis of the displacement of the template grip 140
3.5.2.1 Substrate induced conformational changes of the template grip 141
3.5.2.2 Inhibitor induced conformational changes of the template grip 141
3.5.3.Conformational changes for Leu74-Asp76-Arg78 147
3.5.4 Conformational changes for Glu89-Val90-Glu91-Leu92 149
3.5.5 Conformational changes for Gln151 151
3.5.6 Activation and inhibition of the template grip 152
3.6 The RNase H active site 155
3.6.1 The RNase H active site and hydrolysis of RNA 155

3.6.2 Synopsis of the displacement of the RNase H active site 159
3.6.2.1 Substrate induced conformational changes of the RNase H active site 162
3.6.2.2 Inhibitor induced conformational changes of the RNase H active site 163
3.6.3 Displacement of aspartic acid 443 164
3.6.4 Displacement of glutamic acid 478 166
3.6.5 Displacement of aspartic acid 498 167
3.6.6 Displacement of histidine 539 168
3.6.7 Displacement of aspartic acid 549 170
3.6.8 Activation and inhibition of the RNase H active site 171
3.7 The non-nucleoside inhibitor binding pocket 173
3.7.1 Inhibitor binding 173
3.7.2 Synopsis of the displacement of the NNIBP 175
3.7.3 Conformational changes for tyrosine 181 and 188 179
3.7.4 Conformational changes for phenylalanine 227 and tryptophan 229 182
3.7.5 Conformational changes for proline236 and glutamine 138 183
3.7.6 The binding of non-nucleoside inhibitors 184
3.8 Overall displacement trends and activity correlation 185
3.8.1 Spatial relationships between structural motifs 186
3.8.2 Activity correlation to structural and conformational changes 190
3.9 Integrated mechanism of activation and inhibition 191
3.9.1 DNA vs RNA 191
3.9.2 Snapshots and conformations 192

Chapter 4 – Integrated Drug Design

4.1 - Computer aided ligand design 196
4.1.1 Ligand-based drug design 197
4.1.1.1 Quantitative structure activity relationships 197
4.1.1.2 Comparative molecular field analysis (CoMFA) 199
4.1.1.3 Pharmacophores 201
4.1.2 Structure-based drug design 202
4.1.2.1 Ligand docking 202
4.1.2.2 De novo ligand design 203
4.1.3 Integration of ligand- and structure-based drug design techniques 205
4.1.3.1 Strategy for the integration of ligand- and structure-based pharmacophores 206
4.2 Ligand-based pharmacophore generation with Catalyst 208
4.2.1 Training set selection 209
4.2.2 Conformational analysis 212
4.2.3 Hypothesis generation 213
4.2.4 Hypothesis analysis and selection 215
4.2.4.1 Merged hypotheses 222
4.2.5 Hypothesis validation 224
4.2.5.1 Merged hypotheses 227
4.2.6 Database searching 231
4.3 Structure-based superligand 232
4.3.1 Binding pocket definition 232
4.3.2 Consideration of protein flexibility 236
4.3.3 Superligand generation with Unity queries 240

4.3.4 Database searching with Unity queries 246
4.3.5 Compound selection for biological testing 248
4.3.6 Structural queries in Catalyst 249
4.4 Future directions for integrated drug design 252

Chapter 5 - Conclusions

5.1 Synthetic studies 254
5.2 Conformational studies 256
5.3 Drug design studies 258

Chapter 6 - Experimental

6.1 Synthesis 260
6.2 Computer modelling 283
6.2.1 Modelling of protein conformational changes 283
6.2.2 Ligand-based pharmacophore generation with Catalyst 283
6.2.3 Structure-based pharmacophore generation with Unity 3D 284

References

References 288

Appendices

Appendix 7.1 Subdomains of reverse transcriptase 313
Appendix 7.2 Secondary structure of reverse transcriptase 314
Appendix 7.3 X-ray crystallographic structures of reverse transcriptase 315
Appendix 7.3.1 Additional-ray crystallographic structures of reverse transcriptase 317
Appendix 7.4 Displacement data for comparison of superimposition techniques (I)318
Appendix 7.5 Displacement data for comparison of superimposition techniques (II)319
Appendix 7.6 Subset definitions 320
Appendix 7.7 Displacement data for the PAS 321
Appendix 7.8 Displacement data for the primer grip 323
Appendix 7.9 Distance between the primer grip and the PAS 326
Appendix 7.10 Displacement of the thumb 327
Appendix 7.11 Displacement data for the template grip 331
Appendix 7.12 Displacement data for the RNase H active site 337
Appendix 7.13 Displacement data for the NNIBP 339
Appendix 7.14 Displacement data between motifs of RT 344
Appendix 7.15 Hypoedit utility in Catalyst 345
Appendix 7.15.1 Hypoedit file of activity 'weighted' structural query 346

Chapter 1

Introduction

1.1 The Human Immunodeficiency Virus

The human immunodeficiency virus (HIV) is the causative agent of the acquired immunodeficiency syndrome (AIDS). As the etiological agent of one of the most socially, economically, and medically deleterious epidemics of the late twentieth century it has generated an unprecedented volume of detailed studies of its structure, replication and pathogenesis. Conclusions, conjecture and theories about HIV have not only helped to mitigate its damage, but are applicable to other viruses, and have increased the understanding of multitudinous other cellular and viral systems. A greater understanding of these viral adversaries is needed if a cure or preventative is to be found.

Viruses are biological entities whose genomes consist of nucleic acid. They can only replicate inside host cells using cellular metabolic machinery to form a pool of components which assemble into particles called virions, which serve to protect the genome and transfer it to other cells.[1] Retroviruses, such as HIV, are those viruses whose genome is composed of RNA.

A member of the lentivirus subfamily (Latin *lentus*, slow), HIV is an exogenous, non-oncogenic retrovirus causing chronic diseases of the immune system and the central nervous system. Lentiviruses comprise one of three subfamilies of retroviruses, and usually infect cells of the immune system causing cytopathic effects, such as syncytia, the formation of a large multinucleated cell, and cell death. They contain additional genes to other retroviruses that play a role in efficient virus replication and persistence,

and are unique in their ability to infect non-dividing cells.[1] Lentiviral infections are not cleared by the immune system, leading to accumulated damage over a period of many years and exposing the host to opportunistic infections.[1,2]

There are two predominant strains of HIV. HIV-1, which was isolated in 1983, is found worldwide and is responsible for most cases of HIV infection. HIV-2, isolated in 1986; is endemic in western Africa and is spreading in India. The two strains share a 40-50% homology in their nucleotide sequence, being similar in morphology, ultrastructure, and genomic organisation. HIV-1 is the more virulent and progressive of the two strains, with both strains being typified by chronic opportunistic infections leading to diseases with long incubation periods.

1.1.1 HIV-1 genome and virion structure

The HIV-1 genome (Figure 1.1) consists of two homologous single (+) strands of RNA, 9.8 kb in length.[2,3] It encodes a total of three structural proteins, two envelope proteins, three enzymes, and six accessory proteins. The three coding regions, *gag*, *pol*, and *env* encode the capsid proteins, the viral enzymes necessary for replication, and the external glycoproteins respectively.

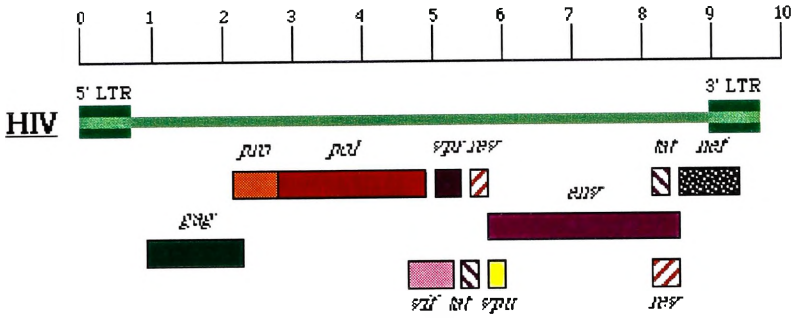


Figure 1.1 HIV-1 genome.

The additional proteins expressed by HIV-1 are either part of the viral particle (*vif* and *vpr*), directly regulate viral gene expression (*tat* and *rev*), or interact with cellular processes to promote virus propagation (*vpu* and *nef*). HIV-2 lacks the *vpu* gene, and instead has a *vpx* gene. Retroviruses such as HIV-1 have the distinguishing characteristic ability to regulate their own gene expression in infected cells. This allows them to establish a long-term association with the host, and generate chronic active infections.

The structural biology of HIV-1 has been intensively studied, and considerable details of its structure and the interactions of its major proteins are available.[2-4] HIV-1 has a cone-shaped capsid core particle, enveloped by a lipid bilayer derived from the membrane of the host cell (Figure1.2).

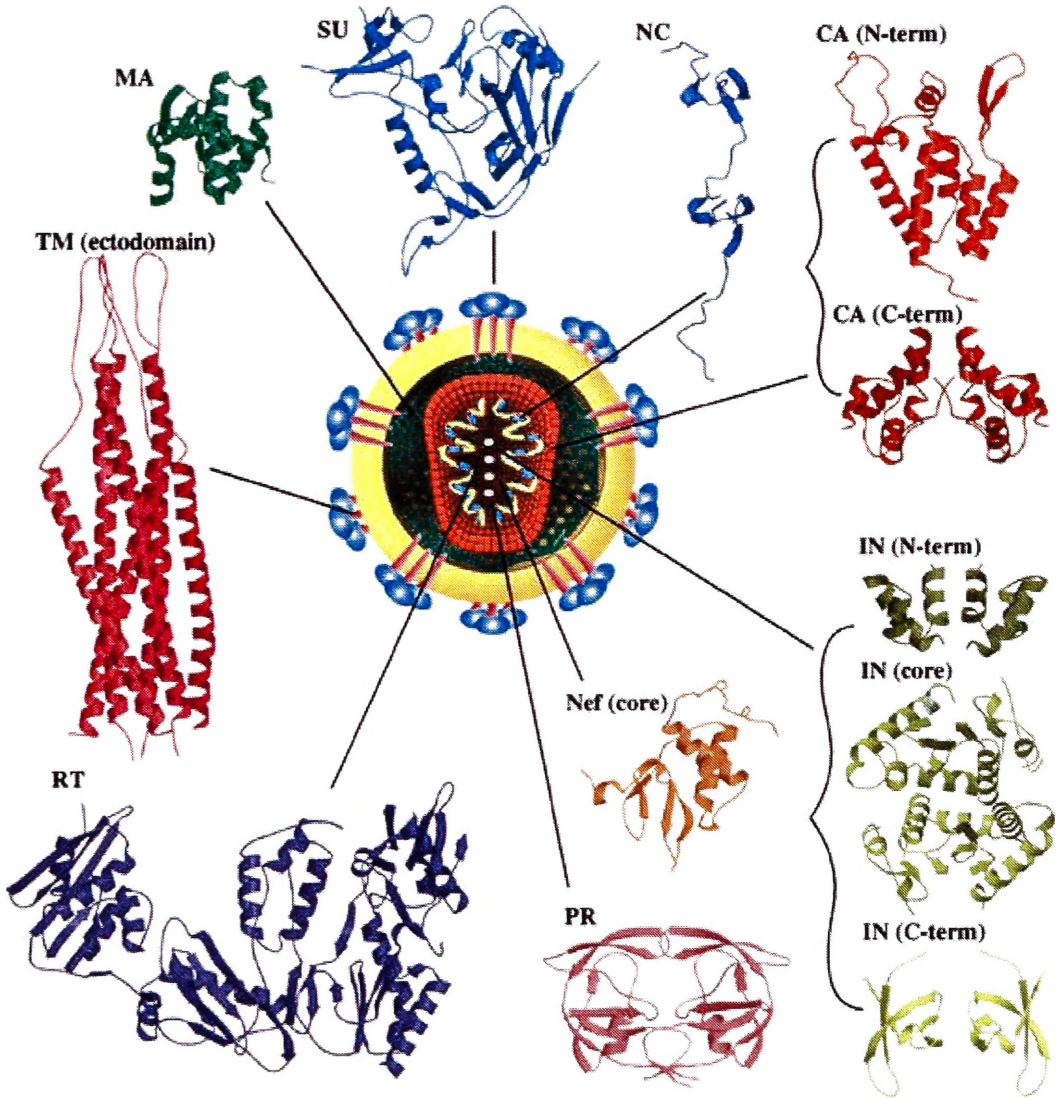


Figure 1.2 HIV-1 structure and proteins.

Exposed surface glycoproteins (SU, gp120) are anchored by interactions with the transmembrane protein (TM, gp41). A matrix shell composed of the matrix protein (MA, p17) lines the inner surface of the viral membrane. At the centre of the virus is a conical capsid core formed by the capsid protein (CA, p24). The core encapsidates two copies of the unspliced viral genome, which is stabilised as a ribonucleoprotein complex with the nucleocapsid protein (NC, p7). The three essential virally encoded enzymes protease (PR), reverse transcriptase (RT), and integrase (IN) are also found in the core.

1.1.2 HIV life cycle

HIV employs a 'Trojan horse' strategy to infiltrate cells of the host immune system, before hijacking its cellular mechanisms and resources to produce more virions.[2,3,5] In productively infective cells, viral infection can cause host cell destruction after 2.2 days.[6-8]

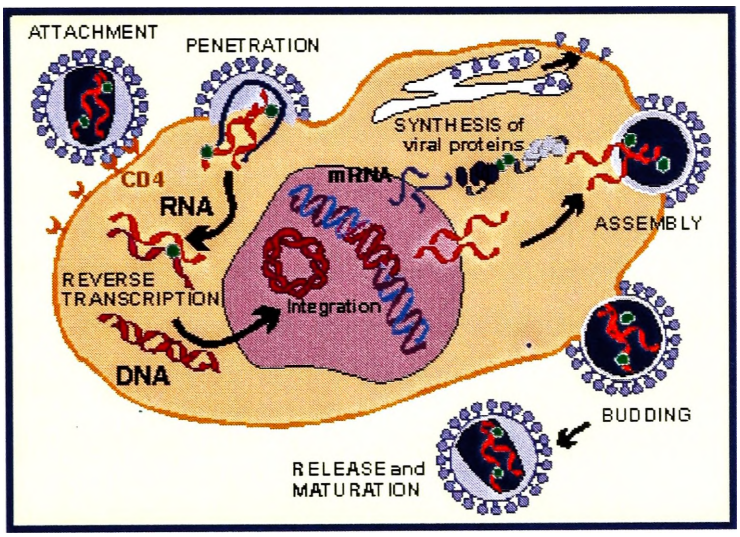


Figure 1.3 HIV life cycle.

HIV-1 binds specifically to cells bearing the CD4 transmembrane glycoprotein, a receptor that normally functions in immune recognition and is found on a variety of cells of hematopoietic origin, such as T lymphocytes, monocytes, macrophages, and B lymphocytes (Figure 1.3). Absorption and fusion of the virus envelope with the host cell membrane is mediated by specific interactions between the viral envelope glycoprotein SU, the amino-terminal immunoglobulin domain of CD4, and additional cell-surface proteins. After penetration, uncoating of the virus core introduces an intracellular reverse transcription complex into the cytoplasm.

Within the first six hours of infection the viral RNA is converted to DNA in the cytoplasm of the cell by a ribonucleoprotein complex containing the genomic RNA together with the NC and MA proteins, and the RT and IN enzymes. The newly synthesised proviral DNA migrates to the nucleus as a nucleoprotein complex where it then integrates with high efficiency into the target DNA, mediated by IN. Integration appears to be random, with the viral DNA remaining permanently associated with the host cell genetic material. Proviral DNA appears in the cytoplasm four hours after infection and can undergo integration after 24 hours.[5]

The proviral DNA can remain in a latent state until cellular activation of the $CD4^+$ T cell by the transcription factor NF-κB binding to promoters in the cellular and viral DNA. Viral gene expression begins with the synthesis of spliced and unspliced messenger RNA (mRNA) transcripts by the cellular RNA polymerase II. The unspliced mRNA is required as the genomic RNA to be incorporated into the virion, and the spliced mRNA transcripts are required for encoding the viral proteins. These are transported out of the nucleus for encapsidation and translation into the proteins of the virus. Transcription and transportation out of the nucleus occur when cellular DNA binding transcriptional factors activate the HIV-1 promoter (*pro*) gene and suitable concentrations of the regulatory *tat*, *rev* and *nev* proteins accumulate. The viral quiescence of some infected primary cells is thus explained since viral activation is induced by cell activation of the infected cell.

The spliced transcripts are translated by endoplasmic reticulum-bound ribosomes before post-translational modification and folding in the endoplasmic reticulum and the Golgi complex. Unspliced mRNA is translated by free ribosomes. Proteolytic processing and mature core formation occur after the virus is released from the infected cell. These stages of the viral life cycle are not well studied or documented.

1.1.3 Immune response to HIV

HIV is carried in infected $CD4^+$ T cells, dendritic cells, monocytes, and macrophages, and as free virus in blood, semen, vaginal fluid, and breast milk. Initial viral infection triggers both humoral and cellular immune responses. Macrophages and related cells engulf free viral particles, digesting them down to small peptides, or epitopes, that are displayed on their cell surface. White blood cells known as helper T lymphocytes ($CD4^+$ T lymphocytes) bear receptors able to recognise a single displayed epitope. Once the helper T cell encounters the right epitope on a macrophage, it binds to the cell, divides and secretes small proteins that help to activate and promote replication of cytotoxic killer T lymphocytes ($CD8^+$ T lymphocytes), B-lymphocytes, and macrophages.

Cytotoxic killer T lymphocytes, once activated, are able to attack and destroy virus-infected cells bearing that specific epitope. Activated B lymphocytes secrete antibodies into the blood and lymph streams that recognise specific epitopes on the viral surface.

The antibodies mark free viral particles, not sequestered in cells, for destruction by other immune cells, such as macrophages.

HIV infection induces cell death, and impairs the immune systems' ability to prime and regenerate its own cells.[9] Cell death occurs primarily via direct viral killing, resulting from the disruption of the cell membrane by massive viral budding and interference with cellular functions by viral proteins. The elimination of infected cells by cytotoxic T lymphocytes that recognise viral epitopes, and the increased susceptibility of infected cells to the induction of apoptosis, or programmed cell death, also contribute to a decline in the $CD4^+$ cell count. Infected cells are also able to fuse with as many as 500 uninfected T cells to form a giant multinucleated cell called a syncytium. A syncytium is able to produce large quantities of virus for a short period of time before dying, within 48 hours of formation. HIV infection is also believed to damage stem cells, thymocyctes, thymic epithelial cells, and lymph nodes responsible for T cell production, proliferation and maturation.[9] Significantly, the infection and depletion of dendritic cells and $CD4^+$ helper T lymphocytes affects the generation, long-term responsiveness and protective function of $CD8^+$ cytotoxic T lymphocytes, the cells most responsible for clearing the virus from the immune system.[10]

1.1.4 Virus reservoirs and the mutability of HIV

A normal aggressive immune response in conjunction with highly active anti-retroviral therapy (HAART) can reduce viral plasma loads below the limits of detection. Yet despite this apparent success, HIV resists complete eradication and progresses to cause the severe immune impairment that defines full-blown AIDS. This evasion and the subsequent devastation of the immune response is due to the virus' ability to reside in latent reservoirs and its propensity to evolve away from any pressures exerted by the host's immune system.

The existence of latent reservoirs provides a seemingly absolute strategy for viral persistence in an infected patient despite potent anti-retroviral therapies.[8,11,12] Sequestering of virus in these reservoirs occurs soon after infection, when there is a period of intense replication and widespread dissemination of virus. Highly active anti-retroviral therapy (HAART) as early as 10 days after the onset of symptoms of primary infection fails to prevent formation of latent reservoirs of virus.[13] Identified reservoirs include extracellular virus trapped as antigen:antibody complexes on the surface of

follicular dentritic cells and lymphoid tissue, in which infected resting memory $CD4^+$ T cells, B lymphocytes, monocytes, macrophages, and dendritic cells are found.[14-17] Although they contribute to less than 1% of the plasma virus, each viral compartment can serve as the 'ember' to re-ignite a high rate of viral replication when a therapeutic regimen is withdrawn,[8,18] or an immune response decays.

Macrophages and dendritic cells are also able to harbor replicating virus without necessarily being killed by it.[17,19] They serve not only as an important reservoir of infection, but are another means of spreading the virus to other tissues such as the brain, testes and retina where most anti-retroviral drugs cannot penetrate. Infected resting and memory $CD4^+$ cells are suspected to have life spans in the order of months, and possibly even years. This stable reservoir may harbour the virus as either unintegrated DNA or in the proviral DNA form, and only requires activation by antigen or host–transcriptional factors to begin production and release of infectious virions. These latent cells provide a long-term memory of viral genotypes no longer present in replicating viral populations.[20] In these reservoirs HIV is active and progressive even when there is little evidence of disease activity in the peripheral blood, and the patient is experiencing clinical latency.[14]

HIV is considered by many to be the most variable virus known to date. The genetic variation, born of remorseless cycles of viral replication, which allows it to finally defeat the immune system, ironically is most probably due to chronic activation of that very immune system.[21] The minimum duration of the HIV-1 life cycle *in vivo* is 1.2 days on average.[8] In that time RT plays a vital role in replication of the viral RNA genome into dsDNA. RT is an extremely error-prone enzyme, lacking the 3'→5' exonucleolytic activity required for proofreading of the growing DNA. All incorrect nucleotides incorporated during (-) strand DNA synthesis from the RNA template are definitively fixed in the viral genome.[22] The mutation rate of HIV-1 RT is in the range of 10^{-3} to 10^{-4} misincorporations per nucleotide per site per replication cycle.[21] Thus mutations occur along the entire length of the genome at the rate of 10^4 to 10^5 times per day.[23] Because of this low fidelity, HIV does not have a fixed genomic sequence, but rather exists as mutant swarms of quasi-species.[21]

In a latent HIV infected patient the average viral production and clearance has been estimated at 1.03×10^{10} virions per day,[8,23] with approximately one half of the circulating virus being replaced with newly produced virions each day.[6,7] In the absence

of any response from the immune system, the viral population would on average double every two days. For a patient who lives with HIV for ten years the viral particles in the body have undergone as much genetic change as humans might experience in the course of millions of years.[20] This enormous diversity is manifested in viral variants bearing fewer recognizable epitopes, making them more difficult for the immune system to detect, and in variants resistant to anti-viral drugs.

The relentless cycles of replication spawn an enormous diversity of viral populations that impair the efficiency of the immune response. Each viral population has different epitopes, but all virions are able to infect any cell of the immune system that bears the CD4 receptor. The cells of the immune system are at a distinct disadvantage as they are designed to accurately respond to a specific antigenic attack, recognizing only selected epitopes; any given immune response is effective against only some viral populations.[10] Once a certain antigenic diversity threshold is reached, the immune system is overwhelmed and becomes totally inefficient in its defense. When the $CD4^+$ T cell count falls below 200 cells.μL^{-1}, from a normal uninfected level of 1200 cells.μL^{-1}; the immune system collapses, and a patient is said to have AIDS.

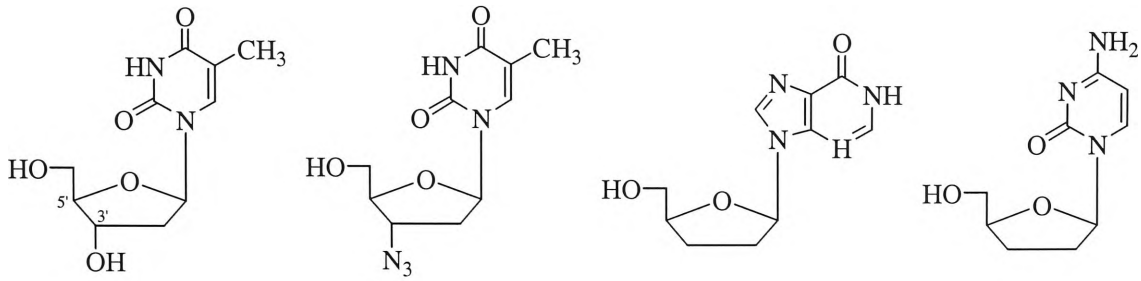
This understanding of the dynamics of the immune system and its response to the continuous replication and evolution of HIV provides the most simple of strategies; by inhibiting or retarding replication, the rate at which mutants arise can be minimised to contain viral diversity. Reducing viral levels and curtailing diversity would help the natural immune system to quell the virus.[20]

1.1.5 Nucleoside reverse transcriptase inhibitors (NRTIs)

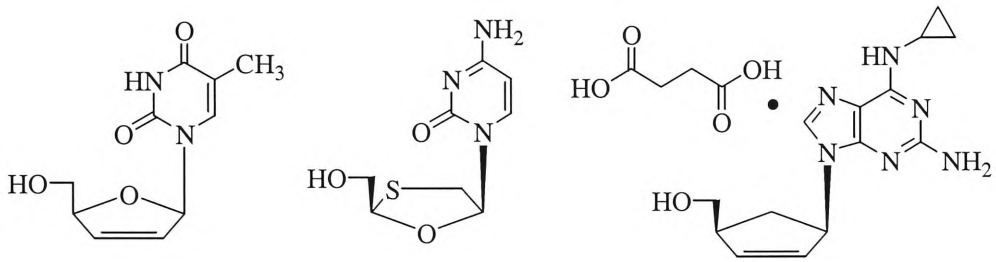
To date no cure has been found for any viral infection. The only alternative course of action is the use of anti-viral drugs to inhibit the virus and bolster the natural immune response to contain infection. Nucleoside analogues were the first and most logical choice for viral inhibition as much of this drug design stratagem had already been studied with anti-cancer drugs. The most attractive viral target is the reverse transcriptase (RT) enzyme, as it is unique to the virus and has no endogenous counterpart in the eukaryotic cell.

Nucleoside reverse transcriptase inhibitors (NRTIs) mimic 2'-deoxynucleotide-5'-triphosphates (dNTPs), naturally occurring building blocks for DNA synthesis, such as thymidine (**1**) shown in its unphosphorylated form (Figure 1.4). They act as

competitive alternative substrates to inhibit the elongation of the growing DNA chain. Once incorporated at the 3'end of the primer strand by RT, they block chain elongation because the 3' position of the ribose is blocked or lacks the OH group necessary for subsequent 5'-3' phosphodiester bonds to form.[24]



Thymidine (**1**) AZT (**2**) ddI (**3**) ddC (**4**)



d4T (**5**) 3TC (**6**) Abacavir (**7**)

Figure 1.4 Nucleoside reverse transcriptase inhibitors.

NRTIs face several disadvantages that undermine their effectiveness as clinical drugs. Despite their selectivity for RT, they can also nevertheless inhibit the activity of normal cellular DNA polymerases α, β, γ and δ leading to myelotoxicity and the inhibition of bone marrow progenitor cells.[25] Their effectiveness is also reduced in metabolically active cells with high pools of naturally occurring dNTPs. NRTIs need to be phosphorylated successively to the 5'-monophosphate, 5'-diphosphate, and 5'-triphosphate form, which can cause problems as some inhibitors have a low affinity for nucleoside kinases that are responsible for the crucial initial phosphorylation. Furthermore the nucleoside kinase activities of some resting cells, such as monocytes and macrophages, may be insufficient to satisfactorily phosphorylate even those analogues that have a high affinity for the enzyme.[24] Additionally the intracellular buildup of partially phosphorylated intermediates may also contribute to toxicity[25]

In 1987, AZT (zidovudine, Retrovir®) (**2**) became the first drug to be approved by the American Food and Drug Administration (FDA) for the treatment of AIDS. It became

the mainstay in the treatment of people with HIV despite its modest and short-term value in slowing the progression of HIV and reducing the number of opportunistic infections. Its success was limited by the rapid emergence of resistant viral strains. ddI (didanosine, Videx®) (**3**) and ddC (zalcitabine, HIVID®) (**4**) were approved by the FDA 1991 and 1992, respectively, for anti-retroviral use in patients with advanced HIV who were intolerant to or deteriorating on AZT. In 1994 d4T (stavudine, Zerit®) (**5**) was approved for use in patients with advanced HIV infection. Unlike ddI and ddC, d4T was found to be effective in penetrating the central nervous system, including the spinal cord and brain; a property which is important for patients who have experienced brain swelling and complications due to toxoplasmosis, or AIDS-related dementia. 3TC (lamivudine, Epivir®) (**6**) was approved for use in combination with AZT in 1995. The addition of 3TC to a therapeutic regimen was found to cause AZT-resistant viral strains to revert to AZT-sensitive strains. The most recent addition to the class of NRTIs approved by the FDA was Abacavir (1592U89 succinate, Ziagen®) (**7**) in 1998. Abacavir requires monophosphorylation and then deamination, before final phosphorylation to the activated triphosphate form. As a monotherapy abacavir appears to have a stronger anti-HIV effect than any of the other approved NRTIs, as well as being less toxic to the hemopoietic system than AZT, but is generally used in combination with AZT, ddI or ddC because of its synergistic anti-HIV activity.

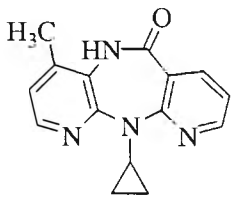
Clinically, the treatment of HIV involves the use of combinations of two, three or even more anti-retroviral drugs.[24] This strategy maximises additive or synergistic affects of different drugs, decreases toxicity by reducing the dosage of the individual compound, reduces the risk of drug resistance development, and can cover a greater spectrum of activity, while achieving potent and sustained suppression of HIV replication. Combination therapy can be successful provided the individual compounds do not have overlapping toxicity profiles, are not antagonistic, can lead to different and preferably mutually antagonising resistance mutations, and are targeted at different viral proteins, or at different active sites within the same viral protein. NRTIs are regularly used in combination with protease inhibitors and non-nucleoside reverse transcriptase inhibitors.

1.1.6 Non-nucleoside reverse transcriptase inhibitors (NNRTIs)

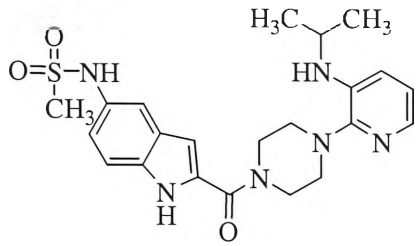
Since the late 1980s, more than 30 structurally different classes of compounds have been identified as non-nucleoside reverse transcriptase inhibitors NNRTIs.[26-28] These allosteric inhibitors are highly specific for HIV-1 RT and have secured a definitive place in the treatment of HIV-1 infections, with three currently approved for clinical use (Figure 1.5) and several others (Figure 1.6) in pre-clinical or clinical development and trials.

NNRTIs are reversible allosteric inhibitors acting at a site distinct from the polymerase active site (PAS) of the RT enzyme, as demonstrated by their non-competitive inhibition with respect to dNTPs and nucleic acid template-primer, and the absence of cross-resistance with NRTIs.[24] NNRTIs have several advantages that make them more efficacious and beneficial than NRTIs. They do not require previous phosphorylation to become active and their non-competitive action with respect to substrates means their action is not compromised in metabolically active cells with high pool levels of dNTPs. Their unique allosteric activity means they do not interact with normal cellular DNA polymerases α, β, γ and δ. This greater specificity for RT has led to an increase in the valuable ratio between the cytotoxic concentration (CC_{50}) required to reduce viability of host cells by 50% and the effective concentration (EC_{50}) required to inhibit HIV-1 induced cytopathicity by 50%, termed the selectivity index. Despite their notoriety for the rapid emergence of drug resistance, this can be circumvented if sufficiently high, but nontoxic, concentrations are used to suppress virus replication and prevent the break-though of resistant virus, or if used in combination therapy.[24]

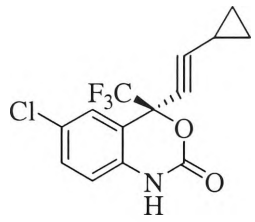
In 1996, Nevirapine (BI-RG-587, Viramune®) (**8**) became the first NNRTI to be approved for use in combination with NRTIs for patients experiencing clinical and immunological deterioration. A member of the dipyridodiazepinone class of NNRTIs nevirapine is a potent anti-retroviral inhibitor capable of crossing the blood brain barrier. Delavirdine (U-90152S, Rescriptor®) (**9**), a bis(heteroaryl)piperazine (BHAP) derivative was approved for use in 1997, and Efavirenz (L-743,726, DMP-266, Sustiva®) (**10**), a benzoxazinone derivative, was approved in 1998.



Nevirapine (**8**)



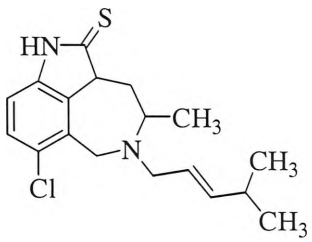
Delavirdine (**9**)



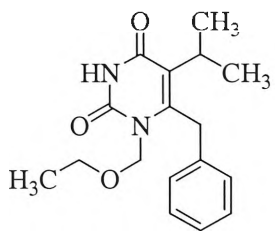
Efavirenz (**10**)

Figure 1.5 Approved NNRTIs.

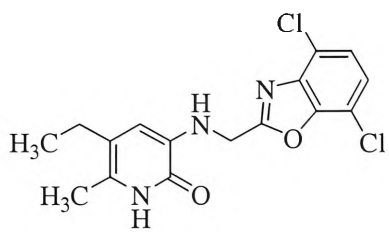
Despite the qualified success of approved NNRTIs with combination therapy in reducing the viral load, the rapid appearance of resistance mutants and the failure of HAART and combination therapy to completely eradicate established HIV infection in patients has assured the continuing need for NNRTIs with greater spectra of inhibition. Those currently in pre-clinical or clinical trial represent just a small assortment of the various classes of NNRTIs (Figure 1.6).



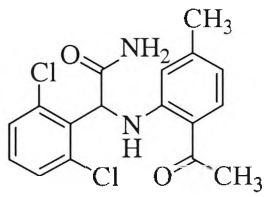
Tivirapine (**11**)



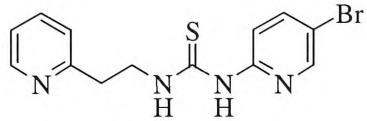
MKC-442 (**12**)



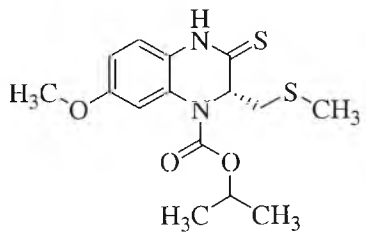
Pyridinone (**13**)



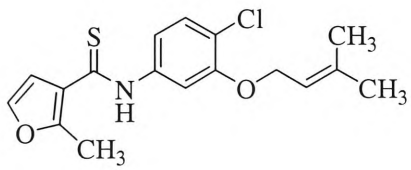
Loviride (**14**)



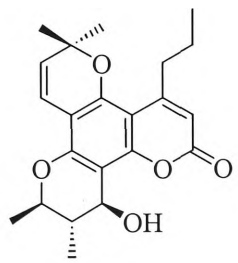
Trovirdine (**15**)



Quinoxaline (**16**)



Thiocarboxanilide (**17**)



Calanolide A (**18**)

Figure 1.6 NNRTIs in clinical trials.

Tivirapine (8-chloro TIBO, R86183) (**11**), an analogue from the tetrahydrobenzodiazepine (TIBO) class, and MKC-442 (I-EBU) (**12**), a hydroxyethoxymethylphenylthiothymine (HEPT) derivative, are actually from those chemical classes first investigated as NNRTIs. Other NNRTIs in current trials include Pyridinone (L-697,661) (**13**), Loviride (α-APA, R89439) (**14**), an α-anilinophenylacetamide (α-APA) derivative, Trovirdine (LY 300046) (**15**), a phenylethylthiazolylthiourea (PETT) analog, Quinoxaline (HBY 097) (**16**), Thiocarboxanilde (UC-781) (**17**), and Calanolide A (**18**), a natural product from *Calophyllum lanigerum*.

All NNRTIs are assumed to bind at the same site, a highly flexible hydrophobic pocket located about 10 Å from the PAS. Although distinct from the PAS, this pocket is spatially and functionally related insofar as the binding of NNRTIs at this pocket results in both short and long-range distortions that disrupt polymerase activity. Short range distortions include a repositioning of the three-stranded β-sheet, in the p66 subunit, containing the catalytic residues required for polymerisation, locking the active catalytic site into an inactive conformation reminiscent of the conformation observed for those same residues in the p51 subunit.[29] The precise geometry and mobility of key aspartic acid residues is distorted such that polymerase activity and its associated reactions no longer function. Longer-range distortions include a hinge movement of the thumb subdomain, which may affect RT activities such as nucleic acid strand transfer, strand displacement, and recognition of primers. These mechanisms are not necessarily mutually exclusive; the binding of a NNRTI may trigger a variety of concerted events resulting in the inactivation or retardation of RT. A greater understanding of how NNRTIs exert these influences would benefit the design and development of improved inhibitors.

Different structural classes of NNRTIs vary with regard to the exact amino acid residues of the pocket site with which they interact, although they all adopt similar butterfly-like shaped conformations.[30] These roughly overlay each other in the binding pocket and are suspected to function as π-electron donors to aromatic side chain residues of the pocket.[31] The corollary to this variation in residue interactions is that a mutation resistant to one inhibitor may not affect another inhibitor; mutations may overlap, only partially overlap, or not overlap at all, as observed with different NNRTIs which do not show cross-resistance to each other. Mutations that have been shown to confer

resistance to NNRTIs cluster around the pocket where the NNRTIs bind suggesting that these resistance mutations lead to a direct alteration of inhibitor binding. Mutations conferring resistance to NNRTIs may affect the shape of the pocket so binding cannot occur, or cause a significant reduction in the favourable binding interactions with the inhibitor. Mutations might increase or decrease the bulk of side chains, leading to steric conflict or loss of contact with an interaction site, or change the local charge or electrical potential of that residue. Mutations may also affect the orientation of significant neighbouring residues that interact directly with the inhibitor. Resistance may involve one, or a combination, of these factors.

Although it was believed that NNRTIs have a high selectivity suppression of HIV-1 replication but little or no inhibitory activity against HIV-2, recent studies[32] have shown that delavirdine, MKC-442, tivirapine, loviride, nevirapine, and pyridinone L-697,661 are able to inhibit HIV-2 strains, although at concentrations 50-fold higher than those required to inhibit HIV-1. Although genetic comparisons of HIV-1 and HIV-2 reveal significant sequence divergence, the amino acid sequence of the important *pol* gene, encoding the RT, PR and IN enzymes, has approximately 60% homology between the two strains.[32] The RT enzyme of HIV-1 and 2 have about the same homology in the residues forming the non-nucleoside inhibitor binding pocket, with the mismatched residues presumably accounting for the differing activities.[33] The similarity in overall structure and functionality of the two enzymes suggests that drugs and drug design strategies effective against HIV-1 RT could also be effective against HIV-2 RT. HIV-2 is the predominant strain in Africa and India, and current HIV-1 specific drugs may require only optimization to fit perfectly the HIV-2 RT non-nucleoside inhibitor binding pocket.

1.1.7 Natural products as antivirals

A common theme that resonates in all areas of anti-retroviral inhibitor research is the incessant need for new and more potent inhibitors with greater chemical structural diversity, and differing mechanisms of inhibition to help delay the development of resistance and control HIV infection. This need can be realised from marine and terrestrial environments, which provide rich sources of novel bioactive compounds to act as lead compounds in the search for clinical anti-HIV candidates. One such example that has reached clinical trials is Calanolide A (**18**) (Figure 1.6), a natural product

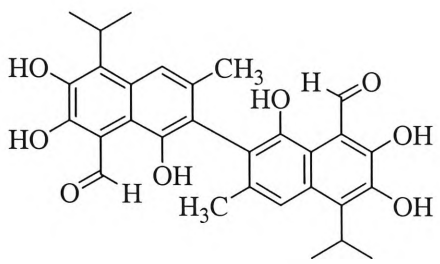
coumarin alkaloid from *Calophyllum lanigerum* var. *austrocoriaceum*, a rainforest tree from Sarawak in Malaysia.[34]

Massive natural product screening efforts from such groups as the US National Cancer Institute (NCI) have brought about the discovery of many new and exciting plant alkaloids that exhibit anti-HIV and anti-RT activity.[35-38] Some of these antiviral natural products are attractive for study since their inhibitory activity has been observed in both strains of HIV, often with similar potencies in both enzyme systems.[36,39,40]

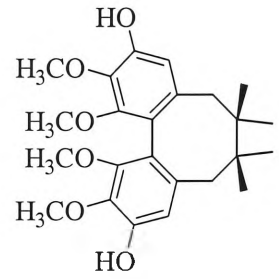
Natural products identified as exhibiting anti-HIV activity span numerous classes of phenolics, terpenoids, tannins, alkaloids, peptides, and carbohydrates. Figure 1.7 illustrates a select few that share an assortment of structural features with Michellamine B (Section 1.2), a novel alkaloid of particular interest.

Gossypol (**19**) is a polyphenolic bissesquiterpene, isolated as a racemic mixture from cottonseed, which has been shown to have selective activity against HIV-1.[41] It exists as (+) and (-) enantiomers, because of restricted rotation of the two naphthyl moieties, although only the (-) enantiomer possess the antiviral activity against HIV and biological activity as an antifertility agent in males. Initial studies appeared to indicate that gossypol had little or no inhibitory activity on HIV-1 RT,[41] but recent testing has confirmed activity in RT assays.[42] Although no information is available for the anti-HIV activity of the gossypol monomer, analogues of gossypol have been synthesised where the dimer has shown activity in inhibiting the replication of HIV, while the monomeric form was ineffective.[43] (-) Gomisin J (**20**) is a dibenzocyclooctadiene-type lignan, isolated from the fruits of *Schisandra chinensis*, found specifically to inhibit HIV-1 RT with an IC_{50} value of 45 $\mu g.ml^{-1}$.[36] A dibrominated derivative, *meta* to the biaryl axis, showed even more potent and selective RT activity (IC_{50} = 1 $\mu g.ml^{-1}$).

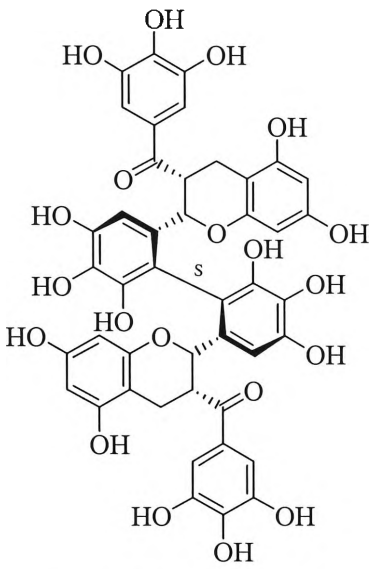
The anti-RT activity of the remainder of these natural products has not been specifically determined, although they all show anti-HIV activity. Theasinensin D (**21**) has shown moderate anti-HIV activity with an EC_{50} value of 8 $\mu g.ml^{-1}$. Interestingly, Theasinensin B which differs only in stereochemistry at the biphenyl bond has no activity.[37,38] Punicalin (**22**) is an ellagitannin which has demonstrated relatively potent inhibitory activity against HIV replication.[44]



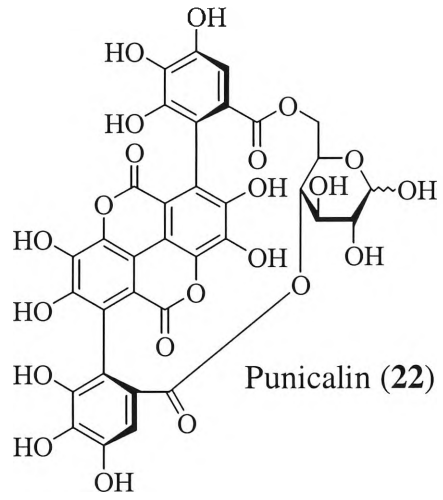
Gossypol (**19**)



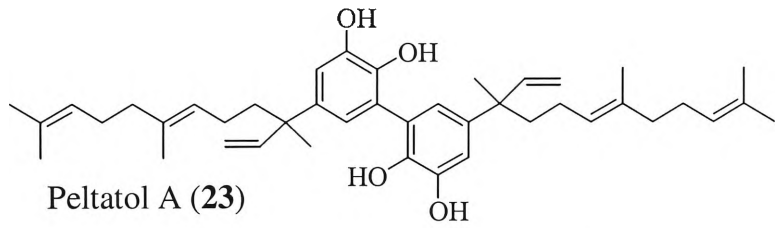
Gomisin J (**20**)



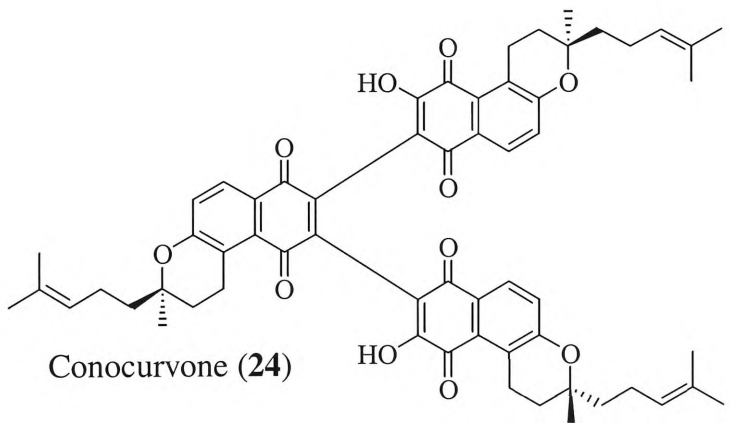
Theasinensin D (**21**)



Punicalin (**22**)



Peltatol A (**23**)



Conocurvone (**24**)

Figure 1.7 Anti-HIV inhibitors from natural sources.

Peltatol A (**23**) and Conocurvone (**24**) are interesting as they were both found to inhibit the cytopathic effects of HIV-1 infection in their dimeric and trimeric forms respectively, but were inactive in their monomeric forms.[45,46] Peltatol A, from the

tropical shrub *Pothomorphe peltata* (Piperaceae), inhibited HIV-1 induced cell killing at subtoxic concentrations of less than 10 $\mu g.ml^{-1}$, yet the monomeric catechol derivative was inactive in the anti-HIV screens.[45] Conocurvone, from the endemic Australian shrub *Conospermum* sp. (Proteaceae), inhibited the cytopathic effects of HIV-1 infection with an antiviral EC_{50} of less than 0.02 μM, with the monomeric unit known as Teretifolione B showing no activity.[46]

These are a few of the natural products (Figure 1.7) that form a loose collective of anti-HIV agents of interest in this study. Although they all posses potent antiviral activity, their precise mechanisms and locations of action are as yet unknown. The salient features of this compilation become more evident after a greater study of principle natural product in this group, the michellamines.

1.2 The Michellamines

In the late 1980s, the National Cancer Institute launched a major initiative to identify novel anti-HIV agents from natural sources. From this massive screening program, the michellamines were discovered in the tropical liana *Ancistrocladus korupensis*, originally misidentified as *A. abbreviatus*, from the Korup rainforest of Cameroon.[39,47] The organic extracts of the aerial components of this liana showed equipotent antiviral cytopathic activity against HIV-1 and HIV-2 virus strains. This was of notable significance since few agents were known to be active against both strains of HIV.[39]

More extensive testing revealed that the michellamines exhibited EC_{50} values (inhibition of viral cell killing) between 1 and 90 μM, and IC_{50} values (cytotoxic concentration) from 40 to greater than 240 μM against a remarkable diversity of strains of both HIV-1 and HIV-2 in a wide spectrum of cell types. Significantly the michellamines provided complete protection against several strains of virus resistant to currently used drugs such as AZT.[47] In murine and canine models, therapeutically significant plasma levels of the michellamines were achieved by intravenous infusion as oral bioavailability proved to be negligible.[48] The michellamines also demonstrated effective inhibition of cellular DNA polymerases α and β, possibly providing an explanation for their narrow *in vitro* therapeutic index in some cell systems.[49]

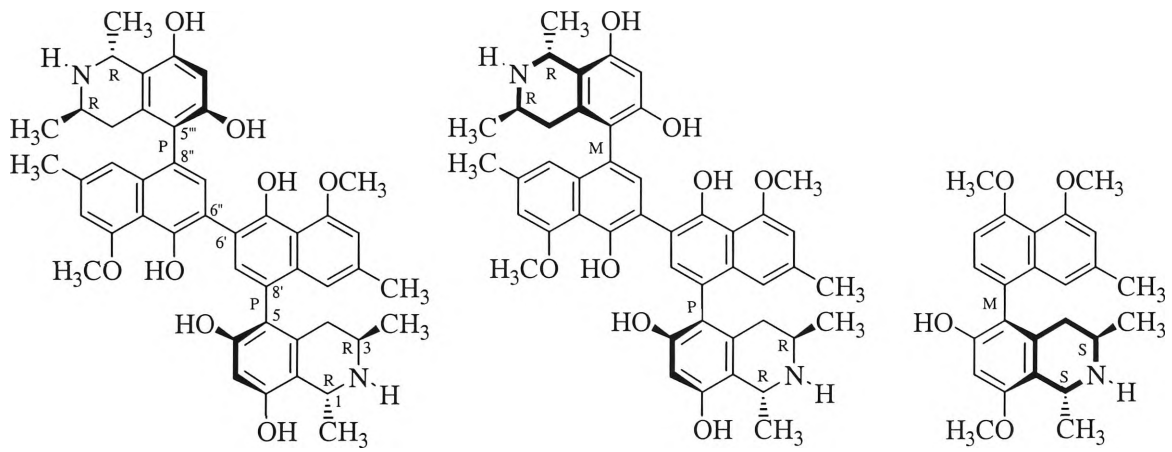
Michellamine B is reported to exert its antiviral activity at two distinct stages of the HIV life cycle. It initially inhibits at the reverse transcriptase stage of the viral life cycle, acting as a non-competitive inhibitor with respect to deoxynucleotide

triphosphates.[49] Additionally it interferes at a later stage by blocking cellular fusion and syncytium formation.[47,49]

Recently the michellamine alkaloids were found to inhibit protein kinase C[50] and demonstrated anti-oxidant activity, as extremely effective free radical scavengers.[51] In antioxidant assays, they were found to be as active as compounds being considered as chemopreventative agents. Aside from their antiviral activity, they could also be of benefit to HIV and cancer sufferers who experience oxidative stress from reactive oxygen species released during the disease process.[51]

1.2.1 Structure of the Michellamines and the Korupensamines

The michellamines were not only of biological and pharmaceutical significance in the search for novel anti-HIV agents. They also represented three new members in an intriguing novel class of naturally occurring dimeric naphthylisoquinoline alkaloids. The michellamines (Figure 1.8) have an unusual C5-C8′ linkage previously only seen in the monomeric naphthylisoquinoline alkaloid Ancistrobrevine B,[39,52,53] found in *A. abbreviatus*.



Michellamine A (**25**) Michellamine B (**26**) Ancistrobrevine B (**27**)

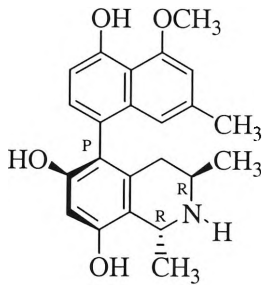
Figure 1.8 Alkaloids from *Ancistrocladus korupensis* and *A. abbreviatus*.

Michellamines A and B were found to be naturally occurring symmetrical coupled products of two atropoisomeric biaryl alkaloids.[54] Atropisomerism arises from hindered rotation about a carbon-carbon bond; one isomer can only be converted to the other by overcoming a sufficiently high-energy barrier.[55] The restricted rotation gives rise to dissymmetric planes, which makes the compound chiral.[56] The rotary characteristics of biaryls are a form of helicity and designated *P* for a right-handed plus helix, when

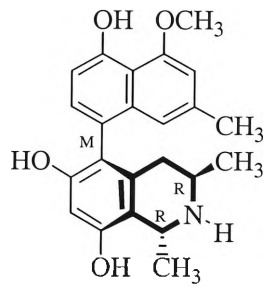
viewed along the axis and moving from the front to the rear, and *M* for a left-handed minus helix.[55,57] Michellamine C was identified as an artefact formed under too harsh isolation conditions.[47] Three more michellamines, Michellamines D-F, were isolated at a later date.[58]

The michellamines are composed of four aromatic systems; two naphthyl and two tetrahydroisoquinoline systems. They have four stereocentres and three biaryl axes; the 6'-6'' axis which is configuratively labile, and the 5-8′ and 8′′-5′′′ axes which are stereogenic due to restricted rotation.[59] They are highly polar with six free phenolic hydroxyl groups and two secondary amino groups. The absolute configuration of Michellamine B, the most naturally abundant alkaloid, was determined to be 1*R*,3*R*,5*R* (*M*), 1′′′*R*,3′′′*R*,5′′′*S* (*P*). Michellamines A and C have the absolute configurations 1*R*,3*R*,5*S* (P), 1′′′*R*,3′′′*R*,5′′′*S* (*P*) and 1*R*,3*R*,5*R* (*M*), 1′′′*R*,3′′′*R*,5′′′*R* (*M*) respectively. Conversion between Michellamines A and C can be achieved by base-catalysed atropisomerisation.[47]

The biogenetic precursors of the michellamines, the monomeric Korupensamines A-D (Figure 1.9) were isolated from *A. korupensis* at a later date.



Korupensamine A (**28**)



Korupensamine B (**29**)

Figure 1.9 Alkaloids from *Ancistrocladus korupensis*.

They showed no anti-HIV activity, but were found to have antimalarial activity against *Plasmodium falciparum* and *Plasmodium berghei*; the IC_{50} values for Korupensamine A were 0.31 and 0.56 $\mu g.mL^{-1}$ for *P. falciparum* and *P. berghei* respectively, and 0.18 and 0.41 $\mu g.mL^{-1}$ for Korupensamine B. Conversely the michellamines showed only weak activity against both of the malaria parasite species, recording IC_{50} values of 20 to greater than 50 $\mu g.mL^{-1}$.[60]

1.2.2 Synthesis of the michellamines

The combined biological, pharmaceutical and structural significance, as well as the scarcity of the natural resources of the michellamines provided a driving force for their total synthesis. Since their discovery, several research groups, using a range of synthetic strategies, have achieved four complete syntheses of the michellamines.

An obvious retrosynthetic analysis of these quadaryls would involve the dissection of the molecule at the three biaryl axes. This presented two obvious synthetic strategies to build the michellamines. The presumable biomimetic pathway would couple the isoquinoline with the naphthalene to form the korupensamine alkaloid, before dimerisation to yield the michellamines (Figure 1.10). The non-biomimetic pathway would involve the formation of the central inner axis in a binaphthyl unit, before coupling of the two outer axes with isoquinoline moieties (Figure 1.11).

The biomimetic pathway was the first synthesis of the michellamines to be realised by the Bringmann group (Figure 1.10).[59] This began with the synthesis of Korupensamines A and B which was achieved by the palladium-catalysed intermolecular biaryl cross-coupling reaction of a trialkylstannalated naphthalene unit (**30**) and a brominated tetrahydroisoquinoline moiety (**31**).[53] A later synthesis of Korupensamine C involved a similar palladium-catalysed cross-coupling reaction of a naphthalene boronic acid moiety and an iodinated tetrahydroisoquinoline unit.[61] Oxidative coupling of a protected derivative of Korupensamine A with silver (I) oxide constructed the central 6′,6′′-binaphthalene bond before reduction and cleavage of the six protective groups yielded the final product (Figure 1.10).[59,62]

The second biomimetic synthesis was undertaken by the Hoye group,[63,64] where the later synthesis of Korupensamine C described previously was used to establish the monomeric unit. The central 6′-6′′ binaphthalene bond was formed by oxidative coupling using silver oxide to give the cross-ring naphthoquinones, which on reduction and deprotection afforded a mixture of Michellamines A-C.

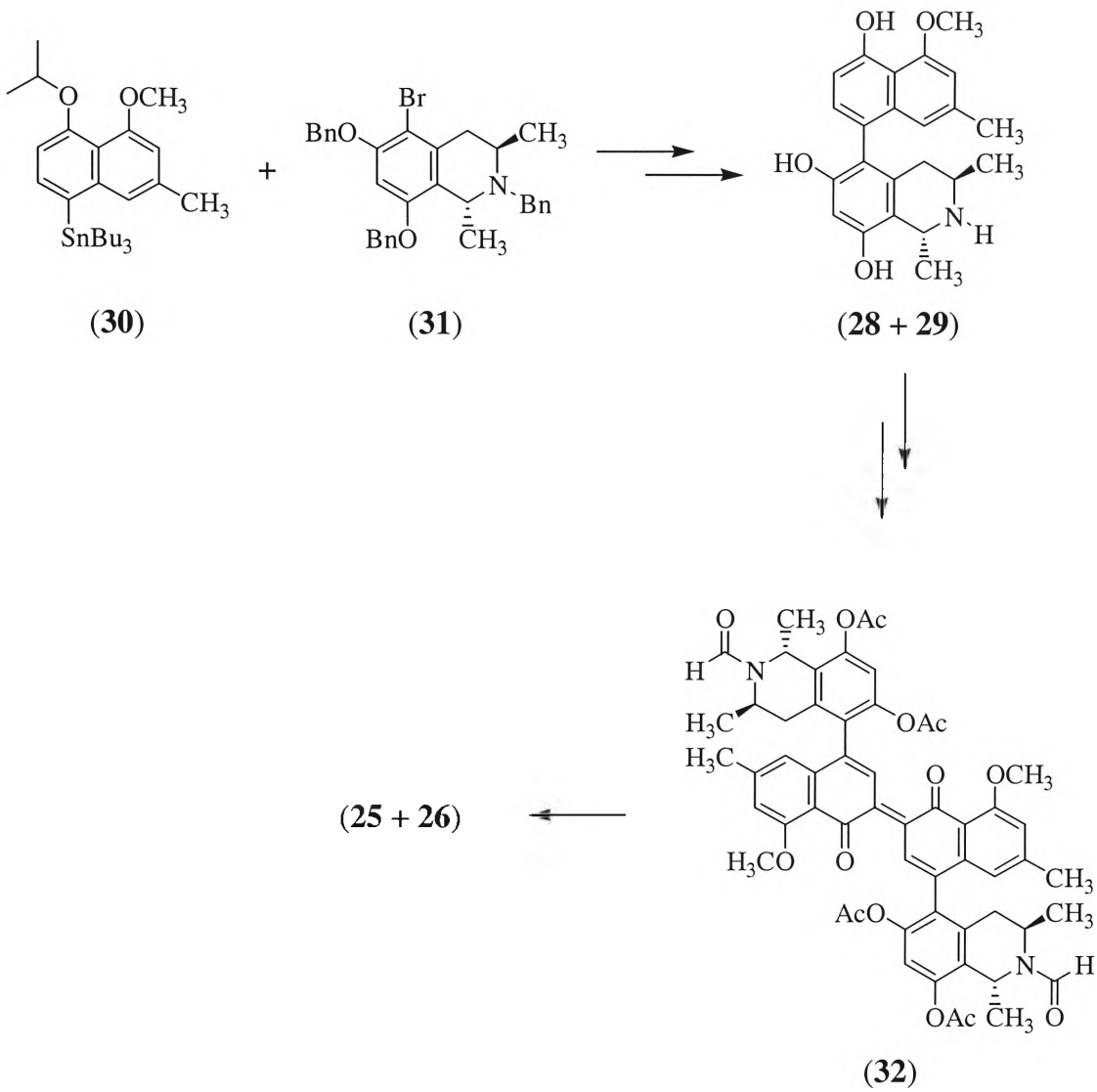


Figure 1.10 Biomimetic synthesis of the michellamines

A third biomimetic synthesis was completed by the Dawson group.[65 66] Similarly to the first two biomimetic syntheses, the korupensamine unit was established with a palladium-catalysed cross-coupling reaction of a naphthalene boronic acid and a brominated isoquinoline unit. In contrast to the previous dimerisation reactions, this synthesis was completed with another palladium-catalysed homo-coupling of a boronic acid derivative of the monomeric unit.

The Bringmann group also undertook the non-biomimetic synthesis (Figure 1.11), where the 6´-6´´ binaphthalene bond was constructed first by using a Pd(0)-copper bronze homo-coupling of a derivatised naphthoquinone (**35**). The bis-*O*-triflate form of the central binaphthalene fragment (**36**) was doubly coupled with a boronic acid isoquinoline building block (**37**), before removal of all protecting groups yielded the final michellamine product.[67,68]

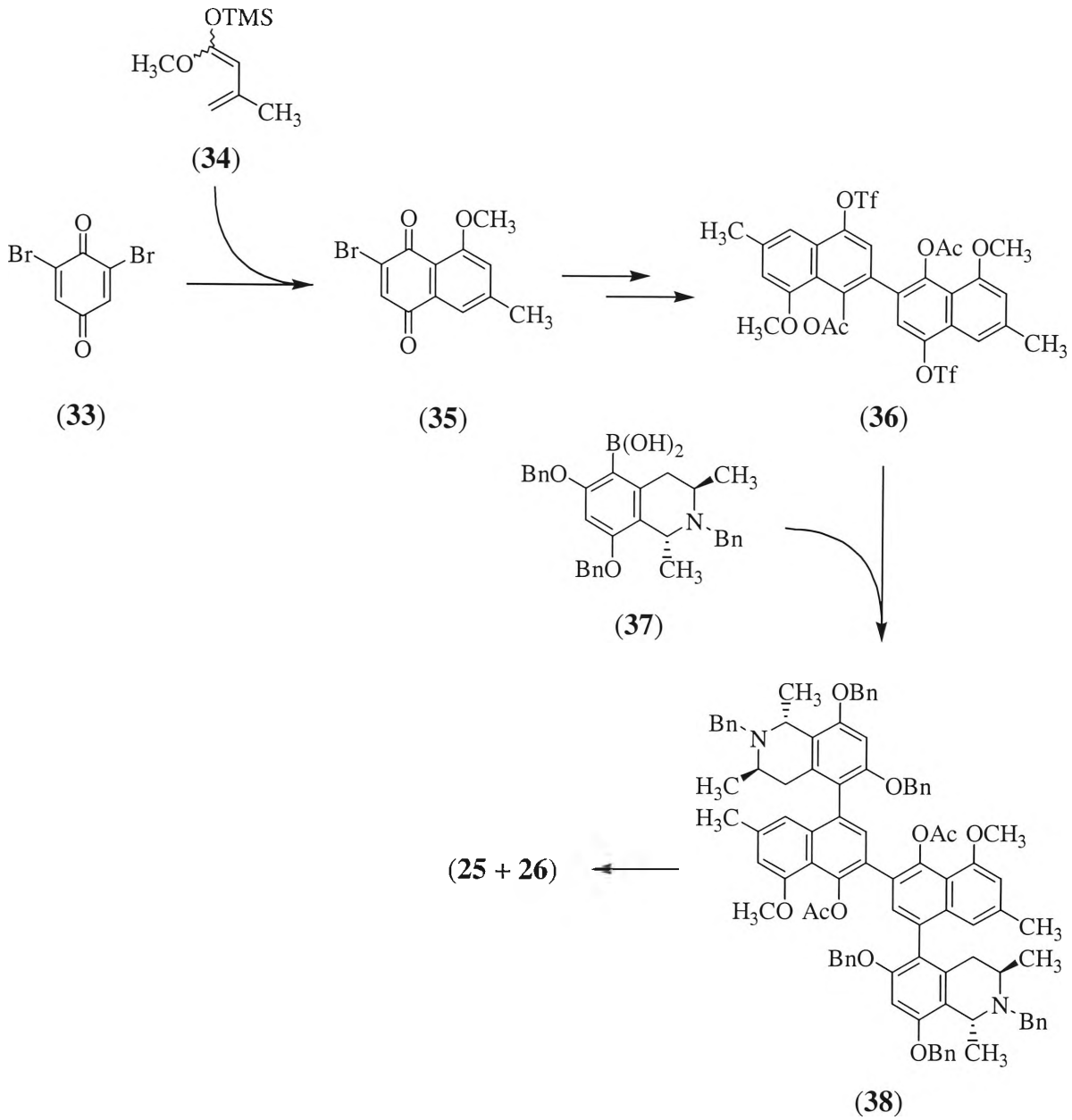


Figure 1.11 Non-biomimetic synthesis of the michellamines

Another more recent strategy involved starting with a biphenyl intermediate (**39**), and the construction of the two 'outer' naphthalene rings by a Wittig and then cyclisation reaction (Figure 1.12).[69] The bromine substituents at the 4,4′-positions in (**43**) allow several options in coupling methodology for the isoquinoline units.

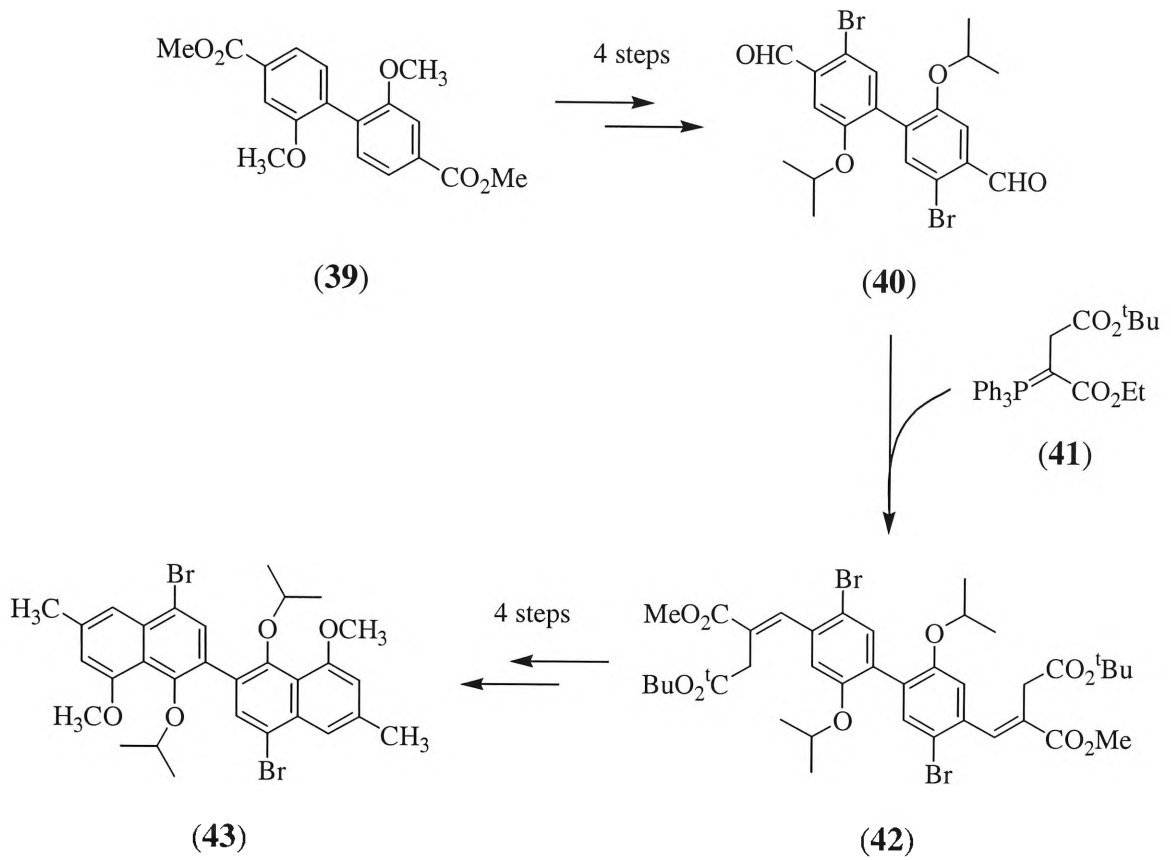


Figure 1.12 Synthetic strategy for the binaphthalene building block representing the central structural element of the michellamines.

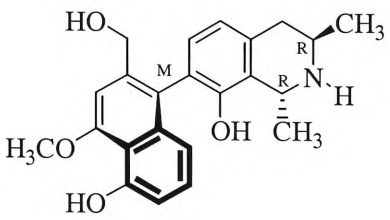
1.2.3 Analogues of the michellamines; antiviral and antimalarial activity

The lack of HIV activity of the korupensamines alluded to the necessity of the binaphthyl structure for anti-viral activity. This seemed to be confirmed by two structural analogues of the michellamines where the central binaphthyl core was modified.[70] One analogue lacked the two methoxy and methyl groups on the central binaphthalene ring system, the other analogue also lacked the methoxy and methyl groups as well as the second tetrahydroisoquinoline aryl group, resulting in an asymmetric compound. These structural changes resulted in decreased HIV RT activity, but interestingly increased protein kinase C activity.[70] The structural significance of the tetrahydroisoquinoline rings was next investigated by the synthesis of a series of structural analogues where those rings were replaced with simpler non-chiral aryl and heteroaryl functional groups.[71] All analogues were found to be devoid of anti-HIV activity up to a concentration of 100 μM, initially suggesting that the heterocyclic portion of the tetrahydroisoquinoline ring was crucial for biological activity. However only anti-HIV activity was investigated; anti-RT activity was not tested for specifically.[71] These analogues were considerably less polar, with several

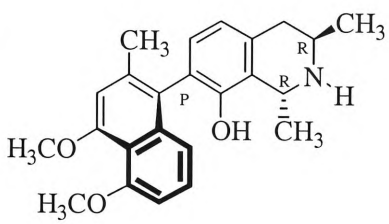
hydrophobic aromatic moieties, and may have had very different transport properties to the michellamines.

Once it was established that the michellamines inhibited the reverse transcriptase enzyme, it was necessary to determine their site of action. Michellamine B was tested on two HIV RT mutants with single amino acid substitutions at positions Tyr181 and Tyr188. These two amino acid residues are part of a hydrophobic non-nucleoside inhibitor binding pocket, and are known to specifically contribute to binding of such inhibitors as nevirapine. Michellamine B was found to inhibit both of the variant RTs with approximately the same potency as that seen with the wild-type HIV-1 and HIV-2 RT, and this was interpreted as michellamine acting at site(s) on the RT enzyme distinct from the other known non-nucleoside inhibitors.[49] This may not necessarily be the case, as other NNRTIs do not always show cross-resistance to all mutations, and further evidence would be required before it was certain that michellamine did not bind in the same binding pocket. Although the size of michellamine may very well be the factor that precludes it from binding at this non-nucleoside site, it also allows for numerously additional and different functional contacts to be made with the enzyme, which may explain the activity retained in the two mutant enzymes.

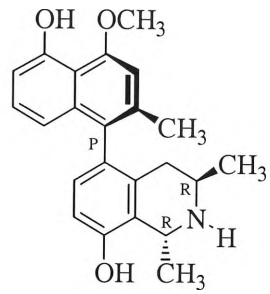
The interesting antiviral and antimalarial activity of the michellamines and korupensamines is seen repeatedly in other related naphthylisoquinoline alkaloids, illustrated in Figure 1.13. Dioncopeltine A (**44**) has demonstrated antimalarial activity against *P. falciparum* and chloroquine-resistant malarial strains,[72] as has Dioncophylline A (**45**) (IC_{50} 1.44 $\mu g.ml^{-1}$, *P. falciparum*),[72] isolated from *Triphyophyllum peltatum* and several *Ancistrocladus* species.[73] The related Dioncophylline C (**46**) also from *T. peltatum* (Dioncophyllaceae) is to date the most active antimalarial naphthylisoquinoline; with *in vitro* activity IC_{50} 0.014 $\mu g.ml^{-1}$ against *Plasmodium falciparum*, it is in the range of medically used drugs.[74,75]



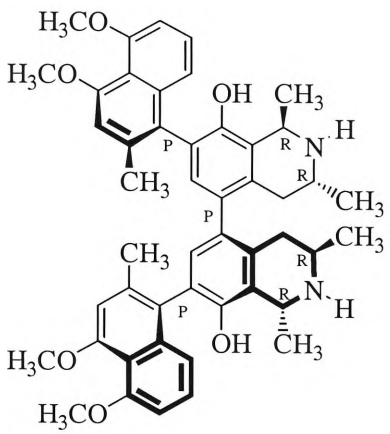
Dioncopeltine A (**44**)



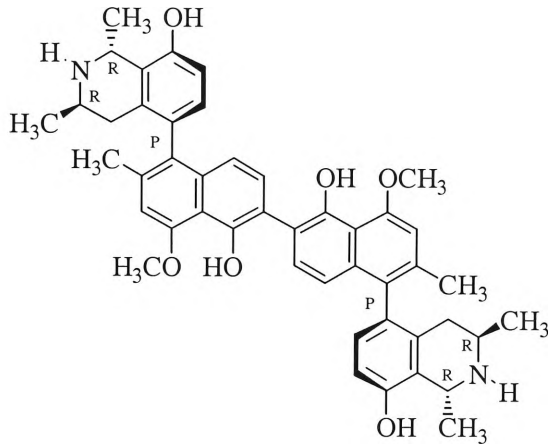
Dioncophylline A (**45**)



Dioncophylline C (**46**)



Jozimine A (**47**)



Jozimine C (**48**)

Figure 1.13 Michellamine related alkaloids with antiviral and antimalarial activity.

Jozimine A (**47**) was the first non-natural dimer described of a naturally occurring monomeric naphthylisoquinoline alkaloid, Dioncophylline A.[76] The antimalarial activity of this dimer, IC_{50} 0.075 $\mu g.ml^{-1}$, is approximately 20 times higher than that of the monomer, a trend seen in several other naphthylisoquinoline systems.[75] This antimalarial activity was unexpected, as the michellamines were entirely inactive against *P. falciparum*; just as unexpected was the lack of any anti-HIV activity.[76] The dimer of Dioncophylline C, Jozimine C (**48**), on the other hand, has less antimalarial activity (IC_{50} 0.445 $\mu g.ml^{-1}$) than the monomer, but has a similar anti-HIV activity to Michellamine B (EC_{50} 27 $\mu g.ml^{-1}$ compared to Michellamine B EC_{50} 14 $\mu g.ml^{-1}$ in the same test system).

1.3 Computational chemistry of biological and chemical systems

Computational chemistry has existed for half a century, and although its exact genesis and original implementation has been a source of ongoing debate, its definition is relatively simple; computational chemistry is the modelling or calculation of any aspect of chemistry by computational means rather than experimental. Although it derives much of its information from experimental sources, computational tools have become

significant mainstays in diverse fields of chemistry, in particular for its application to biological systems. The fundamental rationale for computational chemistry has been a need to put order to observations, and then develop and refine models to explain these observations. This becomes an iterative process to elucidate the workings of basic systems of nature; how a chemical bond forms, how a receptor recognises a substrate, how a primary sequence of amino acid residues folds into a tertiary or quaternary structured protein; the list is truly endless.

Of the challenges in computational chemistry, perhaps the most demanding, promising and rewarding is that of the rational design of medicinal agents for the treatment of human disease.[77] The most important processes that may be involved towards this goal are illustrated in Figure 1.14. Some are not specifically used for drug design, but rather are utilised as tools to model and understand the biological systems in which these drugs potentially may act.

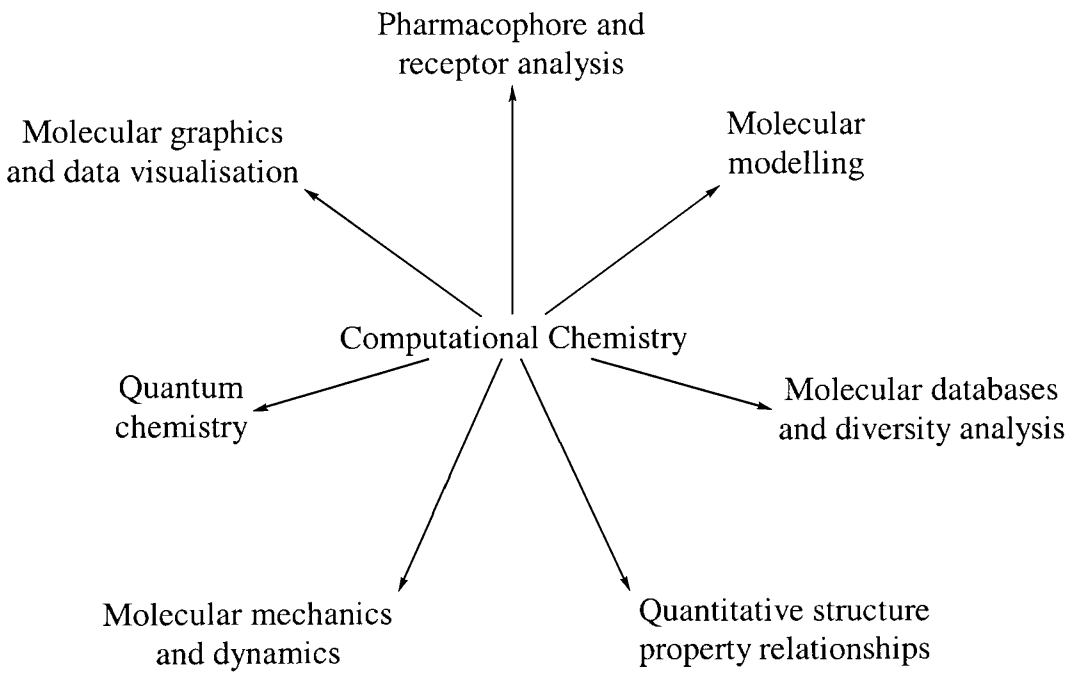


Figure 1.14 Various fields of computational chemistry.[78]

A detailed study of all of these fields of computational chemistry, and the role they play in rational ligand design is beyond the scope of this brief introduction, and have been reviewed and summarised.[78] The areas of most interest in this research were molecular modelling, pharmacophore and receptor analysis, and to a smaller degree, quantitative structure property relationships. Further details of these fields of research are presented in the introductory sections of Chapters 3 and 4.

The roles and contributions of these fields towards the rational design of medicinal agents are illustrated in Figure 1.15.[77]

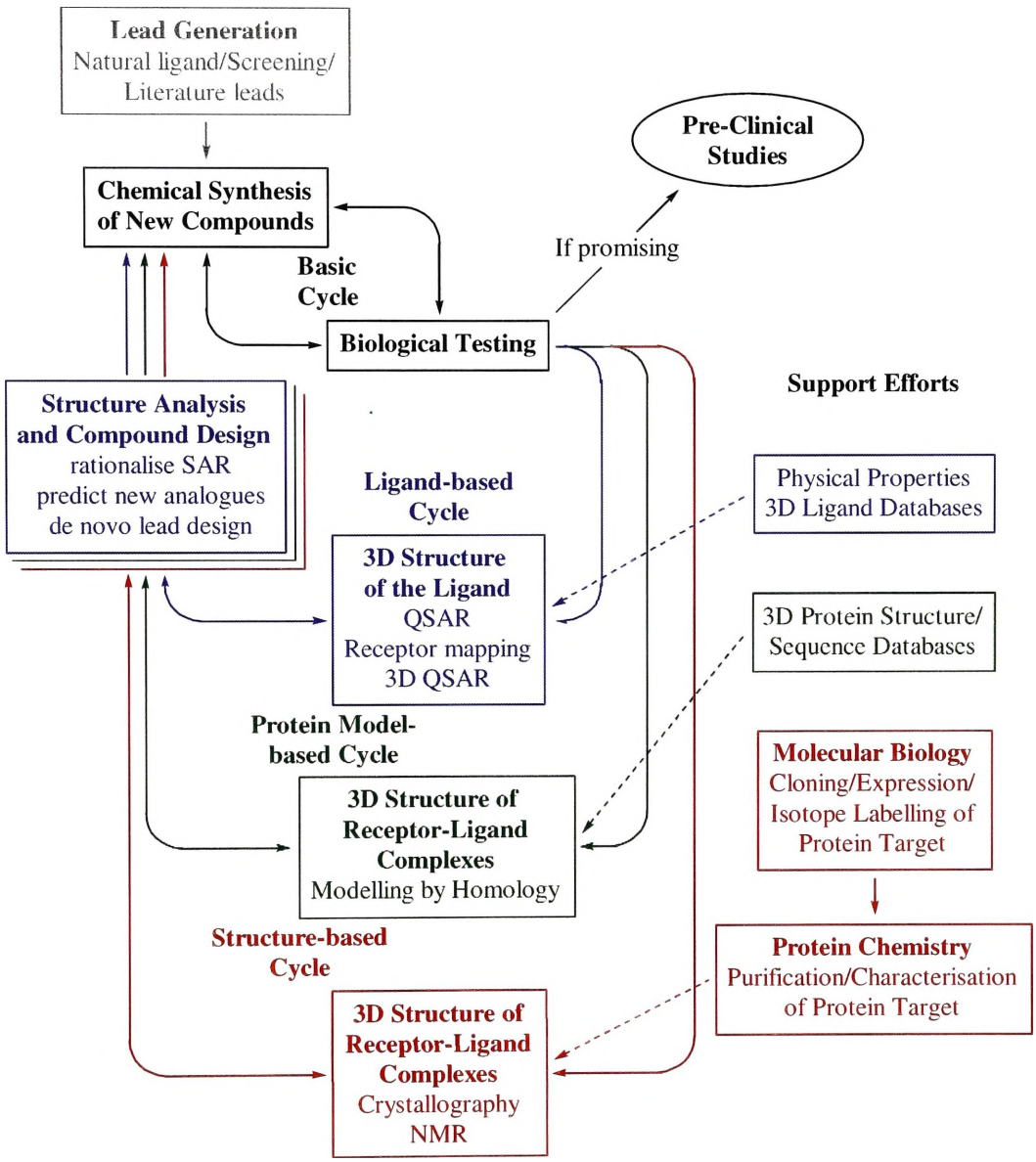


Figure 1.15 Drug design cycles involved in the identification, and optimisation of potential medicinal agents. Results and information from one or more of the three coloured drug design cycles of ligand-based (blue), protein model-based (green), and structure-based (red) drug design feed the basic cycle (black) of biological testing and compound optimisation. The basic cycle continues until a satisfactory compound is produced which can be taken to pre-clinical studies. Each of the drug design cycles of this process, in turn, requires information and results from supporting research. Computational tools are able to manipulate this information, presenting it in a logical and ordered manner to aid in subsequent drug design processes.[77]

The basic, traditional drug design cycle, seen in black, consists of iterative cycles of biological testing for the therapeutic use of the compounds identified by lead generation. This cycle is supplied with new and evolving knowledge developed from other cycles of drug design.

In the absence of information about the structure of the protein, receptor or active site, the blue ligand-based cycle utilises the physical properties of known ligands with biological activity, correlating these properties to their structure to determine the necessary chemical requirements for activity. When the three-dimensional structure of the biological target is known, the computational tools of the red structure-based cycle allow a much more effective study of the compounds considered for synthesis and biological testing, potentially determining the probability of success before the long and tedious process of chemical synthesis is initiated. The green protein-model-based cycle can be considered to be a medley of the blue and red cycles, providing an assortment of data from each, to support the iterative processes of structure determination, analysis, design, synthesis and biological testing.[77]

Beyond the drive to discover and optimise compounds with biological activity can lie the need to understand the interactions of the designed ligands with their receptor protein or active site. Molecular modelling of these biological systems can be invaluable towards greater insights into how these ligands affect their activity. Biological responses may be caused by the ligand blocking the binding site of a natural substrate and suppressing the activity of that enzyme or protein. The ligand may mimic the natural substrate, binding to the active site and inducing the protein to a biological response. In other cases, the ligand may bind to an allosteric site of the protein and influence some other aspect of the protein nature, such as its flexibility, or its ability to dimerise, or form quaternary structures with other proteins or subunits. These effects may be considerable enough to be detrimental to the function of the protein. Molecular modelling of these systems may help to understand how the ligand influences such changes to the protein. If the chemical features that induce these changes can be identified, further study and cycles of the drug design processes of Figure 1.15 may be able to optimise these features to take advantage of their affect on the biological target.

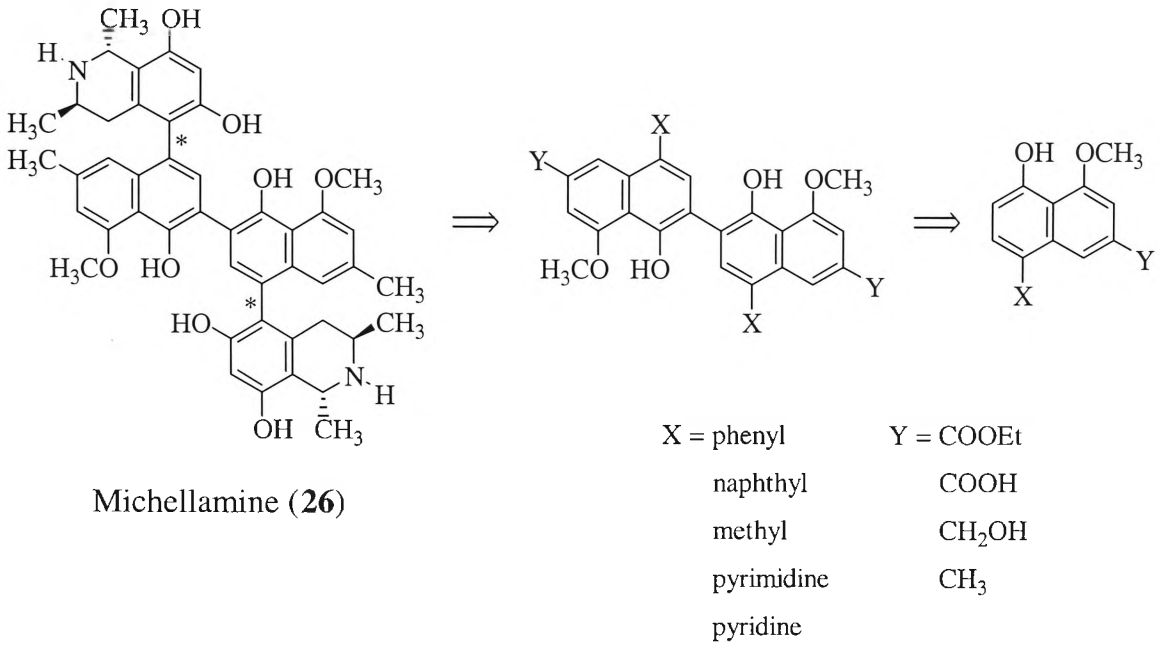
1.4 Research aims

The aim of this research was to study strategies by which to inhibit the life cycle of HIV, in particular, the reverse transcriptase enzyme of the virus. It was proposed to synthesise a series of inhibitors of the reverse transcriptase enzyme based on the naturally occurring michellamine alkaloids. Computer-based modelling studies were designed to learn more about the mechanism by which non-nucleoside inhibitors inhibit the enzyme, as well as about the activation of the enzyme. The third objective of the research was to integrate ligand and structure based drug design techniques to fully utilise all available resources of information for reverse transcriptase inhibitor design.

1.4.1 Synthetic design objectives

The biological activities and structural motifs, which relate together the alkaloids of Figure 1.13 and the natural products of Figure 1.7, provide the basis for a pharmacophore, or template, of antiviral and antimalarial activity. These compounds are all dimeric or trimeric in nature, having some degree of symmetry. They contain a number of aromatic rings with hydrophobic character and they are all highly oxygenated. They have numerous hydroxy groups, particularly *ortho* to the biaryl axes. They all have activity against HIV-1 RT, but for quite a few of them their monomers are inactive compared to the dimeric compound. These monomeric compounds, however, demonstrate significant antimalarial activity.

The objective of the synthetic studies was to incorporate as many of those structural features prevalent in the naturally derived antiviral and antimalarial agents of Figures 1.13 and 1.7 deemed significant for biological activity. The basic structural scaffold chosen was that of the michellamines, as it was the one antiviral with activity against both major strains of HIV. The strategy of Figure 1.16 was employed to design the targeted series of compounds as it also provided access to the corresponding monomers of the same series of compounds. The substituents chosen to replace the tetrahydroisoquinoline moiety of the michellamines varied in size and aromaticity to test steric considerations at this position, and included nitrogen-containing heteroaryls. The methyl substituent on the binaphthyl core was replaced by an ethyl ester functionality (Y) with the intention to progressively transform this to a carboxylic acid, a methyl alcohol and a methyl group to determine the structural requirements at this position.



Michellamine (**26**)

Figure 1.16 Design strategy for the dimeric and monomeric analogues of Michellamine.

1.4.2 Molecular modelling objectives

The HIV-1 reverse transcriptase enzyme, responsible for the reverse transcription of single-stranded viral RNA into double stranded proviral DNA, plays an essential role in the life cycle of the virus, as illustrated in Figure 1.3. Non-nucleoside reverse transcriptase inhibitors are known to allosterically inhibit the replicative role of this enzyme and reduce the infectivity of the virus. The precise basis of this inhibition has not been definitively established, though several hypotheses have been proposed to clarify the mechanism, or mechanisms, of inhibition and rationalise the features of the NNRTIs that induce this response of the enzyme.[79] The three dominant inhibition mechanisms essentially concern the flexibility of the enzyme, and the displacement of several significant structural motifs within the protein. These mechanisms are detailed further in Chapter 3.

The purpose of this research was to develop an objective, and quantitative method to analyse the conformational changes of two active sites, and two nucleic acid grip motifs, as well as the actual allosteric site of inhibition of the enzyme. Using x-ray crystallographic structures of RT in the unliganded form, complexed with double stranded DNA, with and without a nucleotide about to be incorporated, and a series of structurally diverse inhibitors, both inhibitor and substrate induced conformational changes of the enzyme were investigated. The results of this study were intended to either provide support to one or more of the proposed mechanisms of inhibition, or help

conceive a more integrated clarification of the observed biological effect of the binding of substrates and NNRTIs to RT. Additionally, these studies were proposed to characterise flexibility of the allosteric binding pocket and investigate the structural changes induced by the chemical features of the bound inhibitors. The results of these analyses were anticipated to provide further insight into the basic prerequisites for biological activity, and aid in the design of new inhibitors to instigate the same desired inhibitory effects on the enzyme.

1.4.3 Drug design objectives

The abundance of data in the area of reverse transcriptase inhibitory agents, and information about the three dimensional structure of the enzyme provides a rich foundation upon which to base the design of new anti-viral inhibitors. As described in Figure 1.15, the two most established cycles of drug design, ligand-based and structure-based, are segregated by the sources of information used in the design principles and computational tools. This separation is, more often than not, due to necessity rather than choice, as the three-dimensional structures of many receptors and proteins are unavailable, and only structure-activity data is obtainable. This is, without a doubt, not the case with an enzyme like HIV-1 reverse transcriptase. In view of this propitious situation, a drug design strategy was proposed to exploit the wealth of ligand and enzyme structural information.

This research proposed to use ligand-based design techniques to derive pharmacophore models from inhibitors known to bind at the allosteric binding pocket of the enzyme. Furthermore, the inhibitors selected were those that had previously been crystallised with RT, and as such, their bioactive conformations were known. The three-dimensional structures of RT were intended to be used to derive structure-based pharmacophores, describing all potential binding interactions of residues of the binding pocket. The objective of this research was to integrate these pharmacophores and fully exploit the twofold sources of information in one methodology, thereby hopefully combining the strengths from these two design methodologies; and compensating for their shortcomings and limitations.

Chapter 2

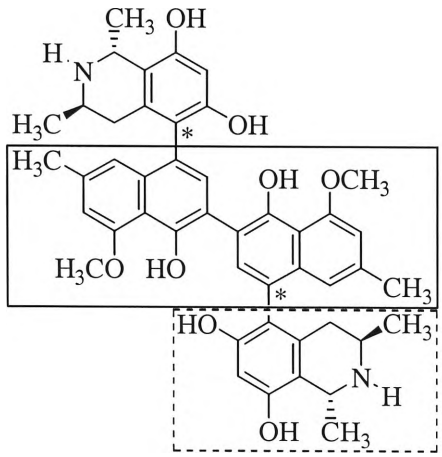
Synthesis

2.1 Introduction

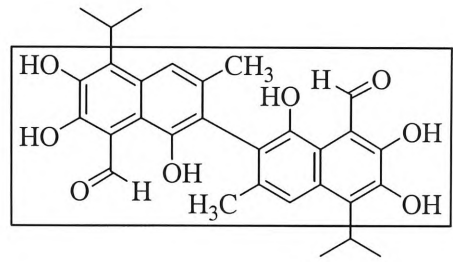
The antiviral activity of the michellamines against both major strains of the HIV virus, and their structural uniqueness as dimeric naphthylisoquinoline alkaloids with an uncommon C5-C8′ biaryl linkage, made them attractive drug targets for synthesis.[59,62-65,67,68] The structural similarity seen between the michellamines (Figure 1.8), michellamine-related alkaloids (Figure 1.13) and various other naturally derived anti-HIV inhibitors (Figure 1.7), such as Gossypol (**19**), provided a lead for the design of potential reverse transcriptase inhibitors. The possibly coincidental antimalarial activity of the michellamine monomers, the korupensamines, and other naphthylisoquinoline alkaloids augmented the interest, and biological and pharmacological scope of the target compounds to be synthesised.

2.1.1 Inhibitor design and strategy

The binaphthyl motif of the michellamines and gossypol, seen boxed in Figure 2.1, was chosen as the basic structural scaffold of the targeted inhibitors, as the michellamines were the principal lead for the inhibitor design. The phenol substituent *ortho* to the biaryl axes was retained, as it appeared in most of the structures of Figures 1.7 and 1.13 considered in the inhibitor design. Similarly the methoxy substituent was included in the target compound as it appeared in Gomisin J (**20**) and all of the naphthylisoquinoline alkaloids. The methyl substituent seen on the boxed binaphthyl core of michellamine, in Figure 2.1was replaced with an ethyl ester functionality for the purpose of further transformation to increase the structural range of the designed inhibitors.



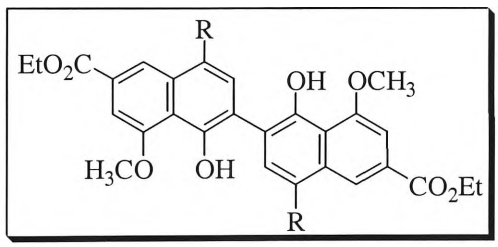
Michellamine (**26**)



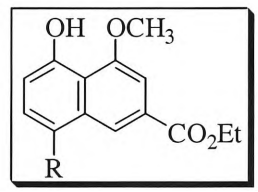
Gossypol (**19**)

Figure 2.1 Binaphthyl motifs of michellamine and gossypol.

The dashed boxed tetrahydroisoquinoline unit of the michellamines was replaced by the simpler substituents (R) of Figure 2.2. Methyl, phenyl, and naphthyl substituents were chosen to provide a succession of sterically different surrogates for the tetrahydroisoquinoline unit. Phenyl and naphthyl substituents were also chosen for their aromatic and electronic features; in the case of the naphthyl substituent, the fused ring system was similar in size to the tetrahydroisoquinoline ring system, but lacked the ring amine. The pyrimidine and pyridine substituents were also chosen for their simplicity and aromaticity, as well as the presence of the nitrogens in the ring system at different positions.



Dimeric Series



Monomeric Series

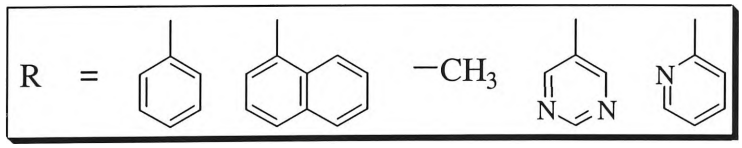


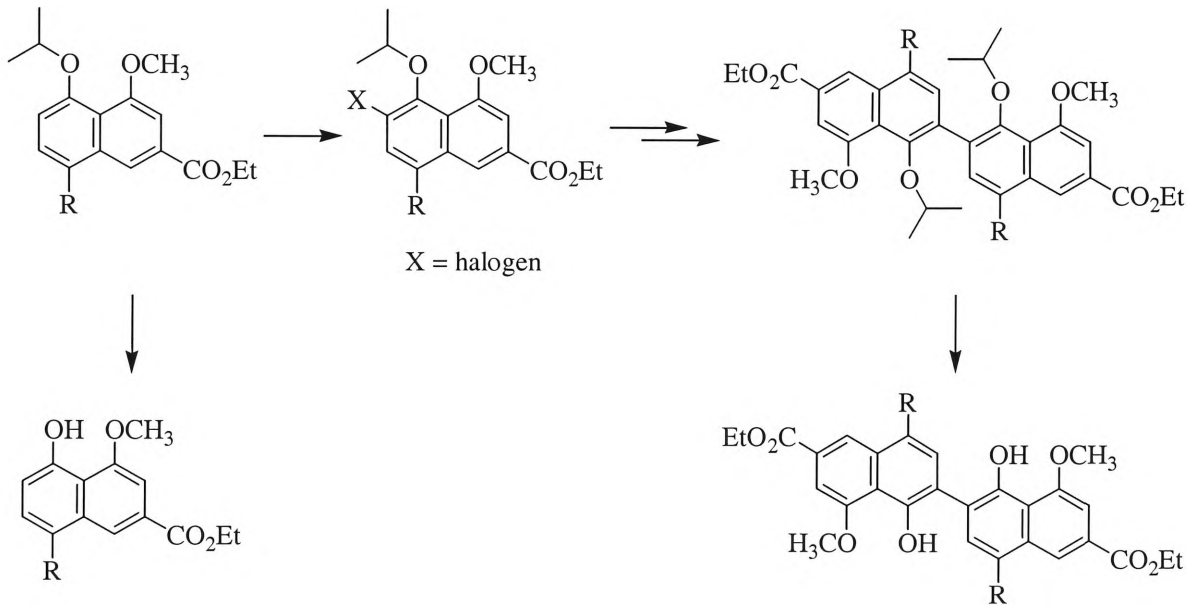
Figure 2.2 Dimeric and monomeric series of designed inhibitors for synthesis.

A dimeric and monomeric series was conceived to investigate the possible requirement of a dimeric structure for antiviral activity, as the dimeric michellamines were active against HIV-1 and 2, while the monomeric korupensamines showed no antiviral activity.[39,47,60] The converse observation that the korupensamines, and several of the

monomeric naphthylisoquinoline alkaloids of Figure 1.13, showed activity against malarial parasites while the michellamines lacked any antimalarial activity provided further motivation and support for the synthesis of the monomeric series. If the synthesised target compounds (Figure 2.2) were to be active as postulated by the inhibitor design, testing results would be expected to show antiviral activity for the dimeric series, and not the monomeric series, and antimalarial activity would be expected in the monomeric series and not in the dimers.

A single convergent synthetic strategy would enable the dimeric series to be completed progressively after the synthesis of the monomeric series, as illustrated in Scheme 2.1. Halogenation at the dimeric biaryl axis position would allow access to several different coupling reactions to complete the synthesis of the dimers. Deprotection of the isopropoxy groups in both monomeric and dimeric series would complete the targeted synthesis.

Scheme 2.1

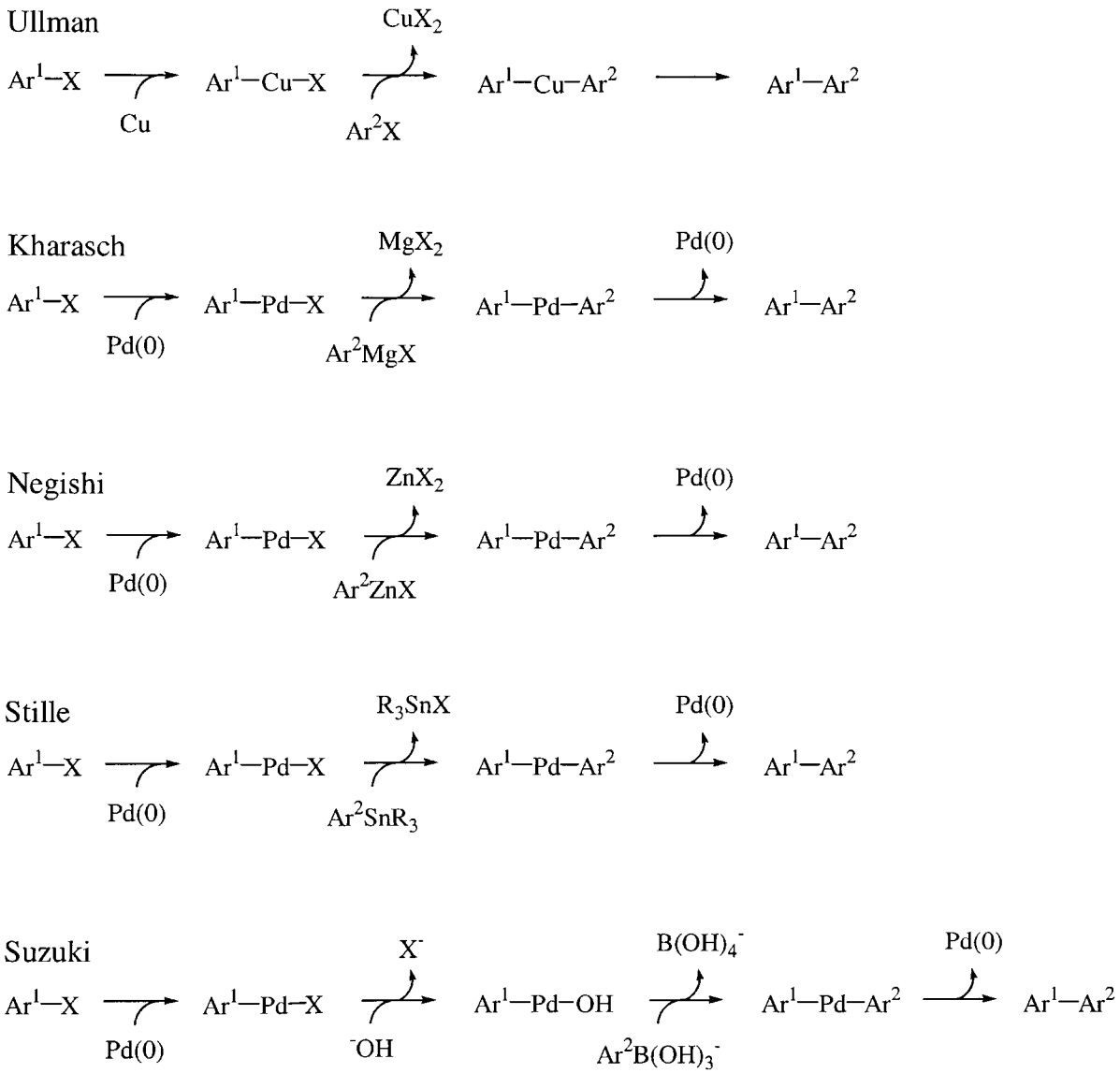


The synthetic strategy adopted was based on two key reactions, comparable to the biomimetic synthesis of the michellamines[59,63] (Section 1.2.2). The first key reaction was a cross-coupling to attach the R-substituent to the naphthyl ring system, and the second was a homo-coupling, illustrated in Scheme 2.1, to dimerise the naphthyl unit and form the binaphthyl target compounds. These coupling reactions to form the biaryl axes proved to be the crux of the synthetic studies.

2.2 Biaryl bond formation

There exist many diverse methodologies for the formation of biaryl carbon-carbon bonds, and foremost among these are the coupling reactions of organometallic species with organic halides and related electrophiles. Scheme 2.2 summarises five of the most common methods for cross-coupling reactions. With the exception of the Ullman coupling, all of these couplings are most commonly catalysed by palladium catalysts.

Scheme 2.2



Ullman synthesis of biaryls comprises reactions in which two aryl halide molecules (Ar^1X, Ar^2X) are condensed in the presence of finely divided copper to form a new aryl-aryl bond with the elimination of copper halide.[80] The reaction is not possible with substituents that provide an alternate path for reaction of the aryl halide, such as amino, hydroxy, and free carboxy groups. Bulky substituents in the *ortho* positions may also

inhibit the reaction by steric hindrance. Furthermore in the synthesis of unsymmetrical biaryls, an optimum yield is obtained only when one of the aryl halides is activated and the other is relatively unreactive.[81]

The Kharasch reaction involves the reaction of an aryl Grignard reagent (Ar^1MgX) with an aryl halide (Ar^2X) in the presence of an appropriate catalyst to yield the biaryl (Ar^1-Ar^2).[80] The Kharasch reaction can also include other functionalised aryls, such as triflates, mesylates, ethers, as well as sulphides and sulphones. Its limitation arises in the polar nature of the Grignard reagent that precludes the use of several types of functional groups in the coupling partner such as aldehydes, ketones, esters and nitro groups.[80]

The Negishi reaction utilises arylzinc reagents (Ar_1ZnX) and aryl halides, triflates, or fluorosulfonates (Ar^2X, Ar^2OTf, Ar^2OSO_2F), and unlike the Kharasch reaction, substituents such as nitrile, ester, and cyano groups can couple.[80,82] The reaction is reasonably able to tolerate substituents *ortho* to the biaryl axis when the coupling partner is an iodine.[83]

The Stille reaction is an extremely versatile route for biaryl synthesis, using arylstannanes (Ar^1SnR_3, R = Me, Bu) and aryl halides or triflates (Ar^2X, Ar^2OTf) as the coupling moieties.[84] The neutral conditions of the reaction mean that many of the substituents not tolerated by the Kharasch and Negishi reactions are often tolerated in the Stille. The flexibility of the Stille reaction has allowed it to be exploited by solid phase synthesis, but its main detraction remains the toxicity of the organotin reagents and by-products.[80]

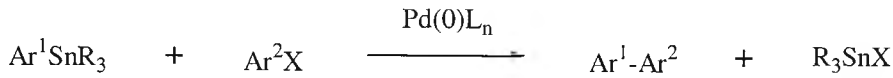
The most recent of the principal biaryl coupling reactions, the Suzuki reaction is extremely versatile as verified by its extensive use in natural product synthesis.[85] It predominantly makes use of boronic acids [$Ar^1B(OH)_2$], although boronate esters and arylboranes are also frequently used, together with aryl halides or triflates (Ar_2X, Ar^2OTf).[86,87] This reaction is especially suited to sterically crowded biaryls.[80]

2.2.1 Stille reaction

In the Stille reaction the biaryl Ar^1-Ar^2 can typically be formed by the coupling of an arylstannane with an electrophile, such as an aryl halide or an aryl triflate, catalysed by palladium[84] (Scheme 2.3). The Stille reaction conditions are compatible with many

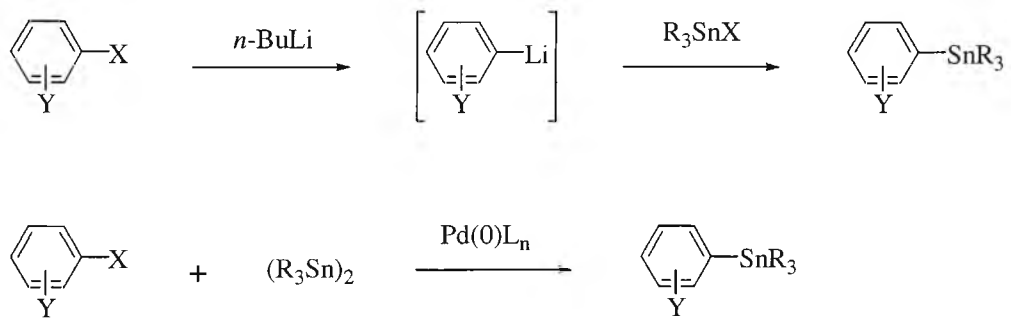
types of functional group such as carboxylic acid, ester, amide, nitro, ether, amine, hydroxy, ketone and even aldehyde groups.[88] As reaction conditions can be very mild, a high degree of stereocomplexity can be tolerated in either of the coupling partners. This makes the Stille reaction particularly useful in natural product synthesis where sensitive functional groups and steroechemistry need to be considered in the synthetic strategy.[89]

Scheme 2.3



The arylstannane can typically be prepared by the reaction of organolithium or organomagnesium derivatives with trialkyltin halides, such as tributyltin chloride or trimethyltin chloride (Scheme 2.4). Another common method of preparation is the palladium-catalysed cross-coupling of hexaalkyldistannanes, usually hexamethylditin or hexabutylditin, with organic electrophiles. Since organostannanes are relatively stable, some can be purified by distillation, and many withstand chromatography on silica gel.[84]

Scheme 2.4

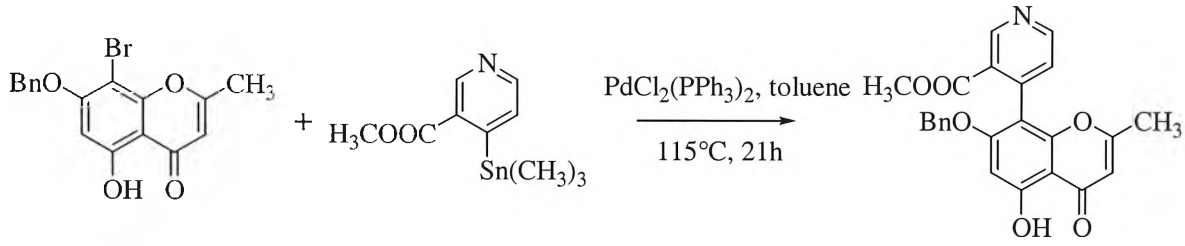


The coupling partner of the arylstannane is usually an aryl halide or an aryl triflate. When triflates are used, an inorganic salt, generally lithium chloride, is usually necessary as a co-reagent.[88] The Stille reaction is often improved by the addition of copper salts,[90,91] the belief is that transmetallation of the arylstannane by the copper salt occurs yielding a more reactive organocopper reagent.

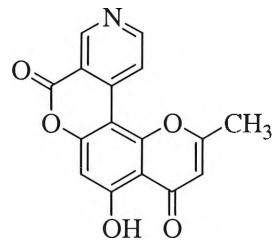
Common by-products in the Stille reaction include the homo-coupled stannane and the homo-coupled electrophile.[84] By-products resulting from the transfer of methyl or butyl groups from the stannane to the coupling partners have been observed in the Stille reaction,[92] although this is usually minor as the order of migration of groups is alkynyl > vinyl > aryl > allyl > benzyl >>> alkyl.[89]

The Stille coupling was used in the synthesis of schumanniophytine (**52**), a lactone alkaloid isolated from *Schumanniophyton magnificum*.[93] This representative of a rare tetracyclic pyranobenzopyranopyridine family of rings systems was prepared by the coupling of a nicotinic acid derivative (**50**) with a brominated chromone benzyl ether (**49**). The presence of the three *ortho* substituents, including two that were electron-donating, made the coupling difficult and accounted for the low yield.[93] Hydrogenolysis of the benzyl group was followed upon heating, by lactonisation to give (**52**)

Scheme 2.5



(**49**) (**50**) (**51**) (25%)

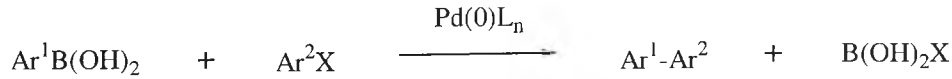


(**52**)

2.2.2 Suzuki reaction

In the Suzuki, reaction the biaryl Ar^1-Ar^2 is formed by the coupling of an arylboronic acid, or an arylboronate, with an electrophile, such as an aryl halide or an aryl triflate, catalysed by palladium[94] (Scheme 2.6). In addition to its flexibility and versatility, with regard to conditions and tolerated substituents, sterically hindered or functionalised biaryls with substituents in the *ortho* positions can be obtained in good yield.[95,96]

Scheme 2.6



Boronic acids do not necessarily need to be synthesised just prior to use as they are more air-stable than stannanes, with only minor quantities of the anhydride (Figure 2.3) found as contaminants after storage.[97] The reaction is largely unaffected by the presence of water, and yields non-toxic by-products.[85] The same functional groups tolerated in the Stille reaction are compatible in the Suzuki reaction. While the Stille reaction occurs in neutral conditions, the Suzuki cross-coupling occurs in basic conditions. The most commonly used base in the Suzuki reaction is Na_2CO_3, but this is often ineffective with sterically demanding coupling reactions, where $Ba(OH)_2$ or K_3PO_4 are used instead for good yields.[85]

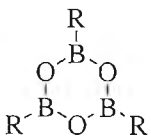
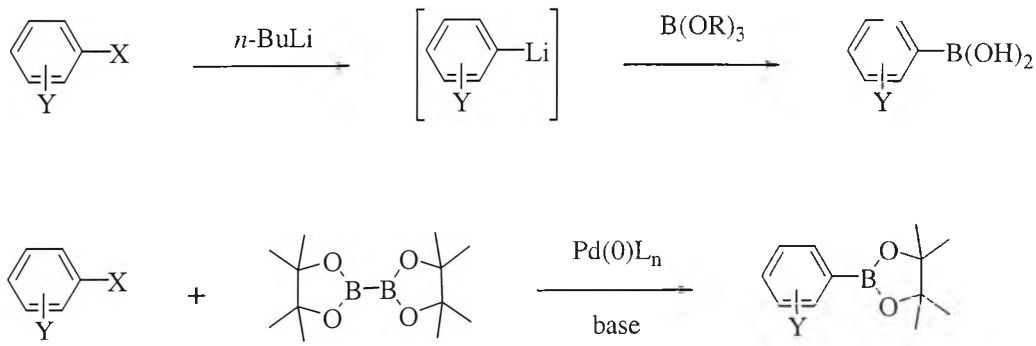


Figure 2.3 Boronic anhydride

Arylboronic acids can be synthesised by converting the corresponding aryl halide into its Grignard reagent, quenching with trimethyl or tributyl borate and aqueous acidic workup.[98,99] Similarly, metal-halide exchange can be achieved with a butyllithium reagent, before reacting with borate reagent and hydrolysis (Scheme 2.7).[100] The introduction of alkoxydiboron reagents have made possible one-pot preparations of unsymmetrical biaryls via a modified *in situ* Suzuki cross-coupling.[101] Scheme 2.7 (bottom) shows the first step in the *in situ* Suzuki cross-coupling reaction where the boronate ester is formed by heating at reflux the aryl halide with the diboron reagent.[101,102] Alkoxydiborons have been shown to be useful boron nucleophiles for cross-coupling with aryl halides, or triflates, that contain base-sensitive functionalities, such as aldehydes, nitriles and esters.[103]

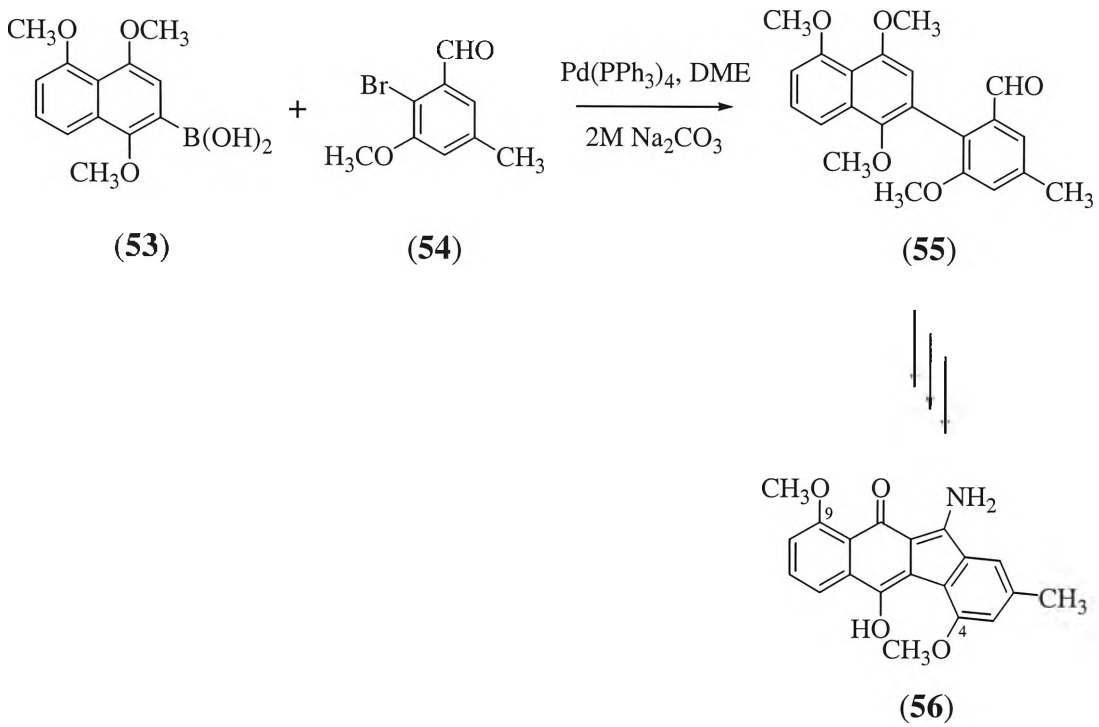
Scheme 2.7



Common byproducts in the Suzuki reaction include the deboronated product,[95] and coupling with a phenyl moiety, usually derived from the catalyst.[104,105] These can usually be minimised by changing the base, catalyst, catalyst quantity, or using phosphine free catalytic systems.[105,106]

The Suzuki coupling was used as a key reaction in the synthesis of the benzo[*b*]fluoren-11-one ring system of $O^{4,9}$dimethylstealthin C (**56**), a methylated derivative of radical scavengers produced by *Stereptomyces viridochromogenes*.[107] The coupled product (**55**) was synthesised in a quantitative yield from the naphthyl boronic acid (**53**) and the bromide (**54**).

Scheme 2.8



(**53**) (**54**) (**55**)

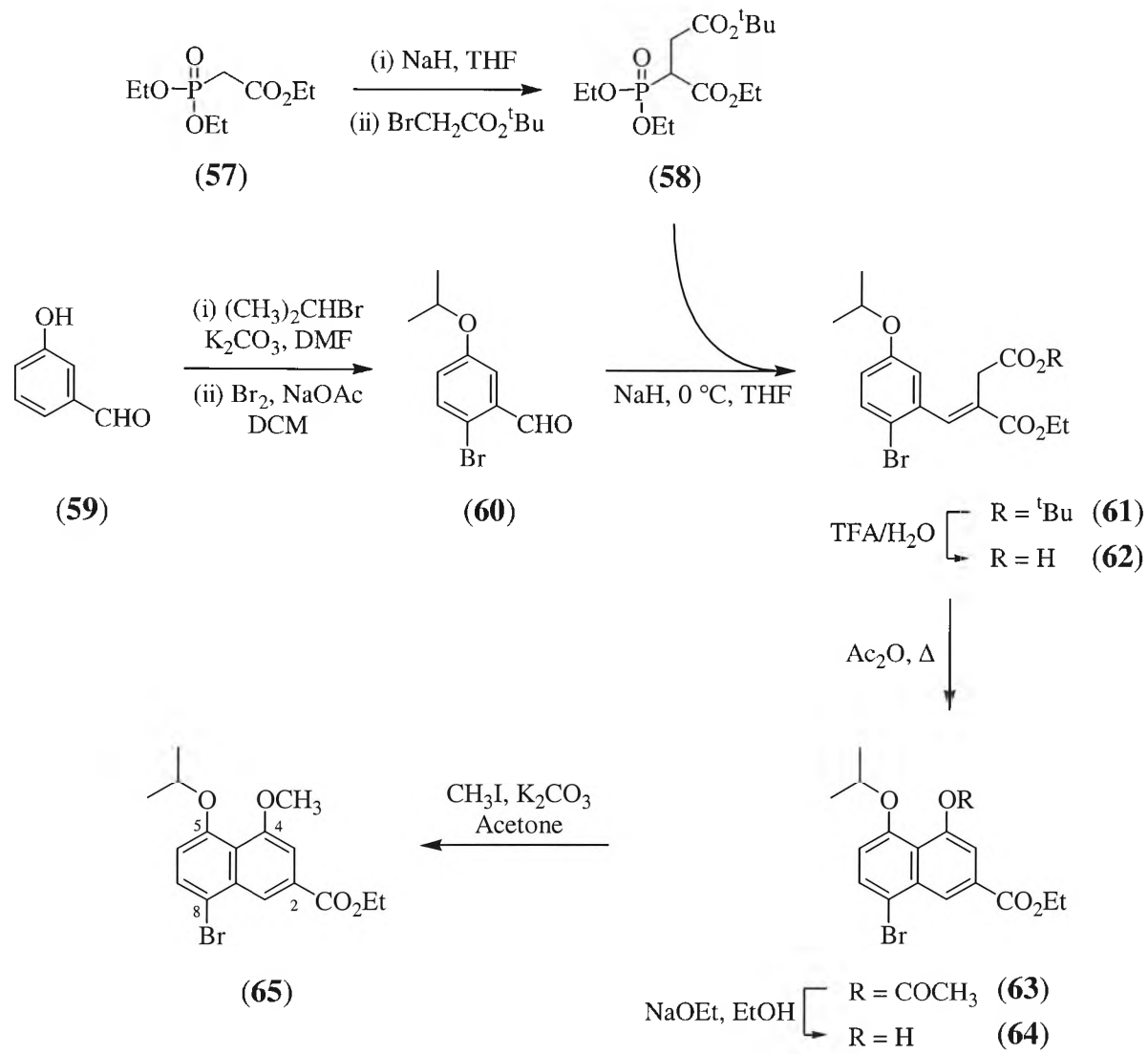
(**56**)

The fundamental mechanism of the Stille and Suzuki cross-coupling reactions occurs *via* three common steps (Scheme 2.2). Initially an oxidative addition of the aryl halide to the catalyst occurs to give an Ar^1-Pd-X complex. This is followed by a transmetallation step, or steps, to give the diarylated palladium moiety, Ar^1-Pd-Ar^2, before a final reductive elimination yields the biaryl product, Ar^1-Ar^2.[80] Generally Pd(II) catalysts are used to promote the cross-coupling reaction, as they are air stable, but require reduction to the Pd(0) state before entering the catalytic cycle.[87,108] The most common catalysts include tetrakis(triphenylphosphine)palladium(0) [$Pd(PPh_3)_4$], bis(triphenylphosphine) palladium(II) chloride [$PdCl_2(PPh_3)_2$], dichloro[1,1'-bis(diphenylphosphino)ferrocene] palladium [$PdCl_2$(dppf)], and palladium acetate [$Pd(OAc)_2$].[80]

2.3 Synthesis of the brominated naphthyl unit (65)

The brominated naphthyl unit (**65**) was synthesised following methodology established for the synthesis of the korupensamines,[53] and the subsequent biomimetic synthesis of the michellamines[62] (Scheme 2.9).

Scheme 2.9



The ethyl ester at the 2-position served as a protecting group for the methyl substituent in the corresponding position in the michellamines, and allowed for other possible transformations, such as to a carboxylic acid or methyl alcohol. The protected hydroxy at the 5-position provided a directing group for the later dimerisation reaction at the 6-position. The bromine handle at the 8-position served as a good leaving group for the cross-coupling reactions to introduce the desired substituents.

The hydroxy substituent of (**59**) was protected by *O*-alkylation with 2-bromopropane to give the isopropoxy derivative in 90 % yield. Halogenation, *para* to the isopropoxy

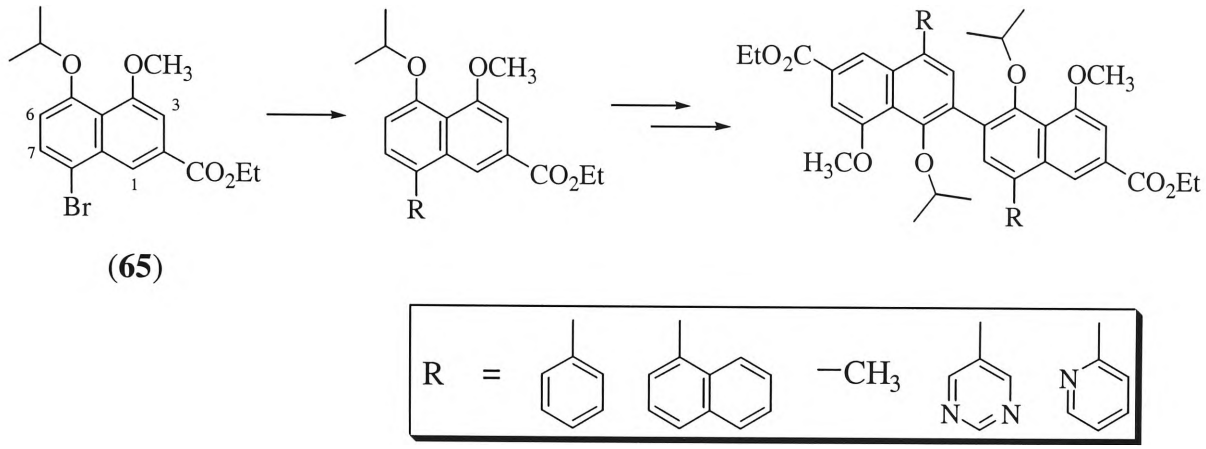
group, with bromine gave (**60**) in a 96 % yield. The C_4-Wittig unit (**58**) was prepared by the addition of sodium hydride to (**57**) followed by the addition of butyl bromoacetate in a 92 % yield.[109] A Wadsworth-Horner-Emmons reaction of the phosphonate carbanion of (**58**) with the carbonyl of (**60**) yielded (**61**) in 56 % yield. Selective hydrolysis of the tertiary butyl group with aqueous trifluoroacetic acid yielded the carboxylic acid (**62**), which was not isolated. The cyclisation of (**62**) with acetic anhydride produced the acetoxy (**63**) and the hydroxy (**64**) naphthyl products, which were not separated. Reaction with sodium ethoxide solution hydrolysed (**63**) to the naphthol (**64**). The yield for the synthesis of (**64**) from (**61**) was 39 %. Methylation of the free hydroxy group of (**64**) with iodomethane gave (**65**) in a 91 % yield. This was used instead of dimethyl sulfate, which gave (**65**) in a 96 % yield, due to safety considerations.

A *M*+1 peak at 367 was identified in the chemical ionisation (CI) mass spectra for (**65**) and *m/z* at 325 (-42) was attributed to the loss of the isopropyl group. The protons of the methoxy group in the 4-position appeared as the singlet at δ 4.00 in the ^{1}H NMR spectrum, while the ethyl ester appeared as a coupled triplet and quartet at δ 1.45 and δ 4.45 respectively; J = 7.1 Hz. The isopropyl methyl groups were equivalent, appearing as a doublet at δ 1.40. The methine hydrogen of the isopropyl group appeared as a characteristic quintet at δ 4.53, instead of the expected septet, due to the smaller peaks being lost in the noise. The naphthyl 1-H and 3-H appeared as doublets at δ 8.56 and δ 7.46 respectively, with a coupling constant of 1.2 Hz, while the 6-H and 7-H also appeared as doublets at δ 6.87 and δ 7.70 respectively, J = 8.4 Hz. Nuclear Overhauser effect spectroscopy (NOESY) identified an nOe between the isopropyl methyl groups and the naphthyl 6-H, expected to disappear after the dimerisation reaction at this position (see Section 2.6).

2.4 Cross-coupling

The synthesis of the monomeric and dimeric series of target compounds was made possible by hetero- or cross coupling reactions of the brominated naphthyl core unit (**65**). The dimeric series required an additional homo-coupling or dimerisation reaction as seen outlined in Scheme 2.10. The precise routes and details for the coupling reactions were determined in part by each targeted series or R group substituent.

Scheme 2.10



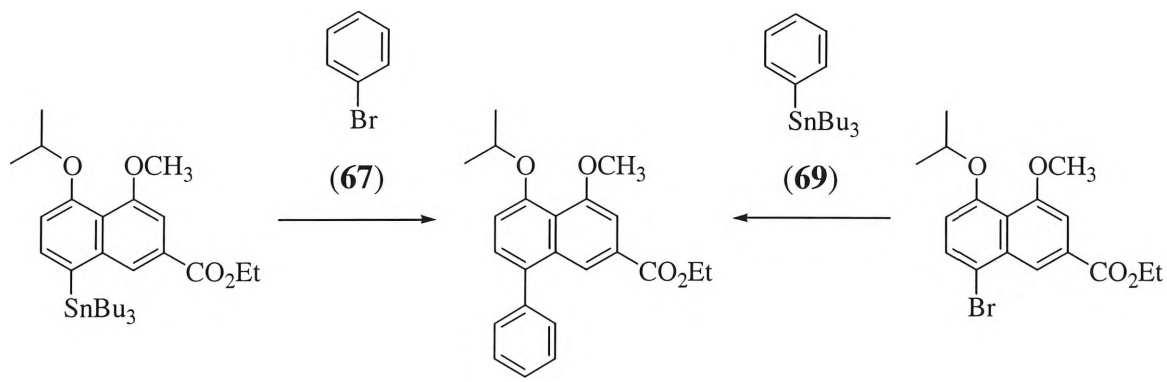
(65)

2.4.1 Stille cross-coupling reactions

The Stille coupling reaction was used in the published syntheses of the michellamines.[62] This provided specific reaction conditions used as a starting point for the cross-coupling experimental procedures.

The bromine handle in the 8-position of the naphthyl core unit (**65**) allowed for two possibilities in the preparation of the coupling units in the Stille reaction (Scheme 2.11). The bromine of (**65**) could be stannylated to the arylstannane and be coupled with the R-substituted aryl and alkyl halides of Scheme 2.10. Alternatively, bromobenzene (**67**) and other aryl halides, could be stannylated to couple with the brominated naphthyl (**65**) to yield the cross-coupled product (**68**).

Scheme 2.11



(66) **(68)** **(65)**

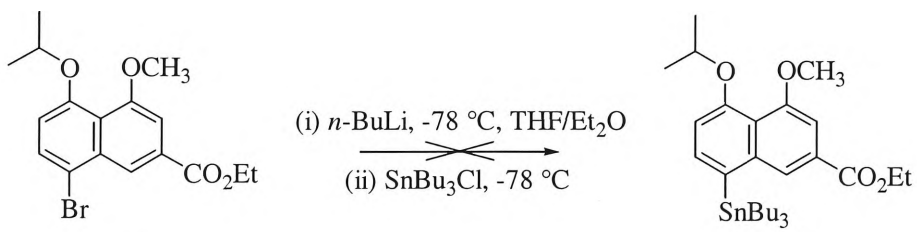
The cross-coupling of the stannylated naphthyl unit (**66**) with aryl halides was the preferred alternative as it had the advantage that once the methodology for the preparation of the naphthyl stannane (**66**) was established, it would not need to be

modified for each different substituent to be coupled at the 8-position. Numerous aryl and alkyl halides are easily available for the coupling reaction. Different R-substituted aryl and alkyl halides could require varying methodologies to prepare the corresponding stannanes and, although commercially available, are best used fresh. Also of note were previous Stille-based syntheses of the michellamines (Section 1.2.2), where the stannane was invariably on the naphthyl, rather than the isoquinoline, unit. The significance of this, and how it might pertain to the nitrogen of the isoquinoline, was uncertain, but may relate to the stannane being preferably placed on the more electronically rich of the two aryl units. As other nitrogen containing heteroaryls were planned to be used as substituents, the same strategy was employed.

2.4.1.1 Preparation of stannanes

Several attempts were made to prepare (**66**) *via* the most common procedure of lithiation of the halide followed by reaction with tributyltin chloride (Scheme 2.12). The reaction conditions were taken from the first biomimetic synthesis of the michellamines,[62] where *tert*-butyllithium was used as the lithiation reagent. In the stannylations of (**65**), *n*-butyllithium was immediately available, and used instead of *tert*-butyllithium.

Scheme 2.12



(**65**) (**66**)

Changes were made to the temperature at the addition of the tributyltin chloride and the solvent composition to induce the synthesis of (**66**) (Table 2.1). A colour change in the reaction solution during the dropwise addition of *n*-butyllithium over ten minutes implied that the first step in the reaction had occurred. However, the addition of tributyltin chloride at –78 °C in entry 1, due to the reagent's potentially volatile nature, yielded no product. The temperature of this addition was carefully increased to 0 °C, in entries 2 and 3, without success. Removing diethyl ether from the solvent composition

had no effect on the reaction result. In several attempts at entries 1-3, approximately 50-75 % of the starting material was recovered, with extraneous polymeric material. Iodine development of all reaction TLCs showed neither the white or yellow stains typical of stannylated compounds. Mass spectra did not show a M+1 at 577 as expected and the ^{1}H NMR was characteristic of (**65**), with no peaks shifting, or additional alkyl peaks. The evidence appeared to indicate that the lithiation reaction failed to displace the bromine atom for the subsequent stannylation reaction to occur.

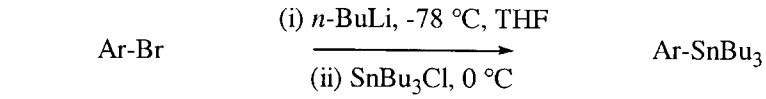
Entry		Conditions	Result
1	(i)	*n*-BuLi, THF/Et_2O, -78 °C	(**65**), polymeric material
	(ii)	$SnBu_3Cl$, -78 °C	
2	(i)	*n*-BuLi, THF/Et_2O, -78 °C	(**65**), polymeric material
	(ii)	$SnBu_3Cl$, 0 °C	
3	(i)	*n*-BuLi, THF, -78 °C	(**65**), polymeric material
	(ii)	$SnBu_3Cl$, 0 °C	

Table 2.1 Conditions used in the attempted synthesis of (66). Entry 1 used 1.0 equivalent of *n*-BuLi; entries 2 and 3 used 1.5 equivalents of *n*-BuLi.

The lack of success in this reaction may have been due to deviations from the literature starting material to be stannylated. The naphthyl unit used by Bringmann *et al.* contained a methyl group at the 2-position instead of the ethyl ester in (**65**).[62] One equivalent of *n*-BuLi may have coordinated to the carbonyl oxygen, preventing the transmetallation with the tributyltin anion. However, this does not explain entries 2 and 3 where 1.5 equivalents of *n*-BuLi were used; some displacement of the bromine by the lithium would have occurred for the subsequent stannylation. The lack of success in the stannylation of (**65**) could not be explained and its synthesis was deferred after the success of the stannylation of R-substituted aryl halide. The use of *tert*-butyllithium could be considered in future attempts at the synthesis of (**66**).

The same procedure of lithiation followed by reaction with tributyltin chloride was used to stannylate the R-substituted aryl halides of Scheme 2.10. Bromobenzene (**67**) and bromonaphthalene (**70**) were used to prepare the corresponding stannanes in Scheme 2.13.

Scheme 2.13



Ar = phenyl (**67**)
naphthyl (**70**)

Ar = phenyl (**69**)
naphthyl (**71**)

Entries 1 and 2 of Table 2.2 showed that bis(tributyltin) and trimethyltin chloride were not successful as stannylating agents, and only starting material was recovered. In the lithium mediated stannylations, arylstannane products (**69**) and (**71**) were only isolated when tributyltin chloride was used (entries 5 and 6). An alternative preparation of stannanes involving reflux with a palladium catalyst and stannylating reagent was attempted with some success in entry 4. The use of bis(tributyltin) and palladium tetrakis(triphenylphosphine) refluxing for 24 hours, in entry 3, yielded no product. The addition of triethylamine, as an organic base, and the substitution of $PdCl_2$ and PPh_3 as the catalyst, generated *in situ*, resulted in a 57 % yield of (**69**). This was less than the yield of 70 % achieved by lithiation and subsequent stannylation of bromobenzene (**67**). All reactions required at least one silica column to remove the excess tin reagents. In retrospect, these conditions could have been used to prepare the naphthyl stannane (**66**), but the low yielding reaction would have been an unfavourable option after the multi-step synthesis required to produce (**65**).

Entry	Ar-Br	Conditions	Result
1	(**67**)	(i) *n*-BuLi, THF, -78 °C (ii) $(SnBu_3)_2$, 0 °C	(**67**)
2	(**67**)	(i) *n*-BuLi, THF, -78 °C (ii) $SnMe_3Cl$, 0 °C	(**67**)
3	(**67**)	$(SnBu_3)_2$, $Pd(PPh_3)_4$, toluene, 115 °C	(**67**)
4	(**67**)	$(SnBu_3)_2$, $PdCl_2$, PPh_3, triethylamine, toluene, 115 °C	(**69**) 57 %
5	(**67**)	(i) *n*-BuLi, THF, -78 °C (ii) $SnBu_3Cl$, 0 °C	(**69**) 70 %
6	(**70**)	(i) *n*-BuLi, THF, -78 °C (ii) $SnBu_3Cl$, 0 °C	(**71**) 76 %

Table 2.2 Conditions used in the synthesis of (69) and (71) Scheme 2.13. 1.5 equivalents of *n*-BuLi and/or stannylating agent were used in each reaction.

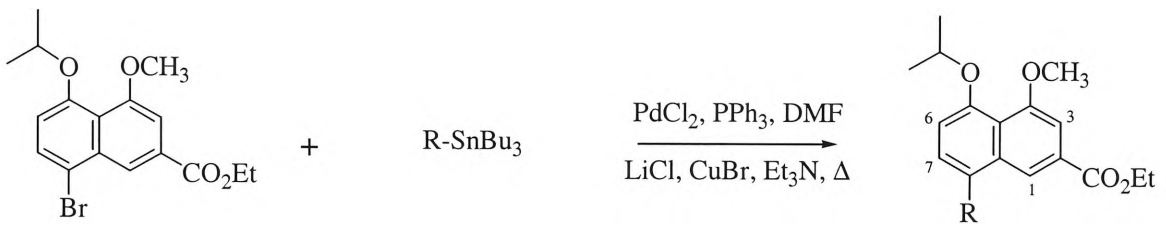
The products (**69**) and (**71**) appeared as white spots after iodine development of reaction TLCs. Tributyltin incorporation was confirmed by M+1 peaks at 367 and 369 in the

mass spectra of (**69**), for the two major isotopes of tin, 118 and 120. The mass spectra of (**71**) showed *M*+1 peaks at 417 and 419.

2.4.1.2 Coupling of naphthyl core halide and aryl stannanes

The conditions for the Stille reaction were also based on those used in the synthesis of the michellamines.[62] Due to the immediate availability of the reagents, palladium chloride and triphenylphosphine were used as the catalyst instead of the palladium chloride bis(triphenylphosphine) and triphenylphosphine Bringmann *et al.* used. Triethylamine was also added to regenerate the catalyst that was formed *in situ*.

Scheme 2.14



(**65**) R = phenyl (**69**), naphthyl (**71**)

R = phenyl (**68**) (15%), naphthyl (**72**) (17%)

In a typical reaction (**65**) was added to palladium chloride, triphenylphosphine, triethylamine, lithium chloride, and copper (I) bromide in DMF, and heated at reflux, under nitrogen, for 24-48 hours. The reaction was monitored and quenched after TLC indicated no trace of (**65**), the rate-limiting reagent. The yields for the reactions, 15% and 17% for (**68**) and (**72**) respectively, were low and extreme anhydrous conditions and the rigorous exclusion of air using freeze thaw cycles did little to increase the product yields. The addition of a few crystals of 2,6-di-*tert*-4-methylphenol as an anti-oxidant[62] affected no change to the reaction yield of (**68**) or (**75**). Apart from the removal of a large amount of polymeric material in the purification process by silica column chromatography no explanation can be made for the up to 85% of starting material (**65**) unaccounted for.

The 1- and 7-H of (**68**) were shielded by the phenyl substituent and were shifted upfield; from δ 8.56 to δ 8.22 and δ 7.70 to δ 7.36, respectively, compared to (**65**). The 3-H signal was virtually unchanged, but 6-H was notably shifted downfield from δ 6.87 to δ

7.09 [§]. The protons of the phenyl substituent were not resolved, and appeared as a multiplet. Nuclear Overhauser effect spectroscopy showed an nOe between the 1-H of (**68**) and the phenyl protons *ortho* to the biaryl axis. Mass spectra showed a *M*+1 at 365.

Similarly the 1-, 3- and 7-H of (**72**) were shielded by the naphthyl substituent and showed upfield shifting to δ 7.74, δ 7.39, and δ 7.41 respectively, while the 6-H was shifted downfield to δ 7.14. ^{13}C NMR revealed 27 carbon resonances as expected. The isopropyl group of (**72**) appeared not to be freely rotating, with the two methyl groups appearing as two separate doublets, in the ^{1}H NMR, compared to the one doublet seen in (**68**). In addition, two separate carbon signals at δ 22.16 and δ 22.22 were observed in the ^{13}C NMR, as if the two methyl groups of the isopropyl were chemically nonequivalent. A temperature series of ^{1}H NMRs were run at 25, 35, and 45 °C to hopefully allow the isopropyl group to rotate freely, and appear as one doublet rather than two. The resulting spectra showed only a very slight decrease in the separation of the doublets.

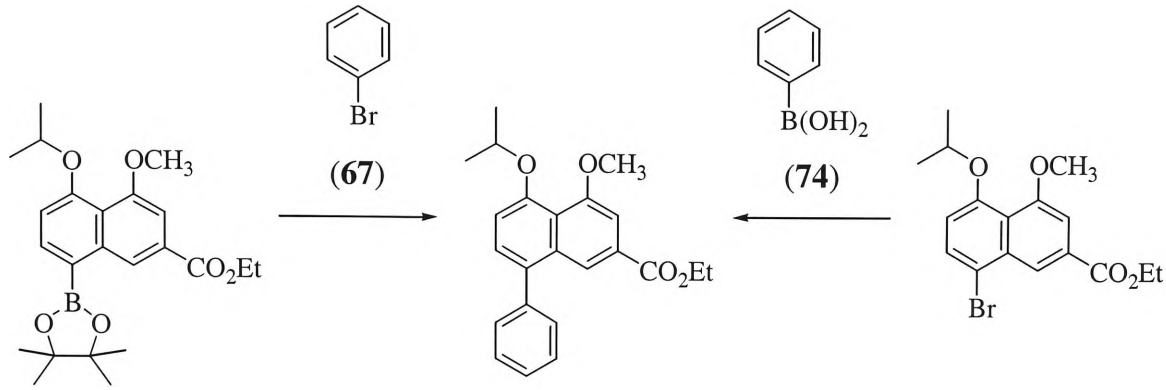
The unacceptable yields of these Stille reactions detracted from these coupling reactions. The additional inability to recover starting material (**65**) from the reaction mixtures made this practically infeasible as a route to the target compounds. Concurrent Suzuki-coupling studies proved to be more successful, and the Stille cross-coupling reactions were terminated at this stage.

2.4.2 Suzuki cross-coupling reactions

Similar to the strategy employed in the Stille cross-coupling reactions (Scheme 2.11), Scheme 2.15 outlines the two possible synthetic routes to (**68**). A high yielding synthesis of (**73**) would allow access to halide or triflate substituents that could not be easily transformed to boronic acids themselves, and would avoid possible modifications for the synthesis of various aryl boronic acids from aryl halides or triflates. Readily available boronic acids, such as phenyl boronic acid (**74**), would also allow the synthesis of the target compounds from the brominated naphthyl (**65**).

[§] The naphthyl shifts in the cross-coupled products were most apparent in the pyrimidine substituent, where the aromatic hydrogens were more clearly resolved, Figure 2.5 Section 2.4.2.3.

Scheme 2.15

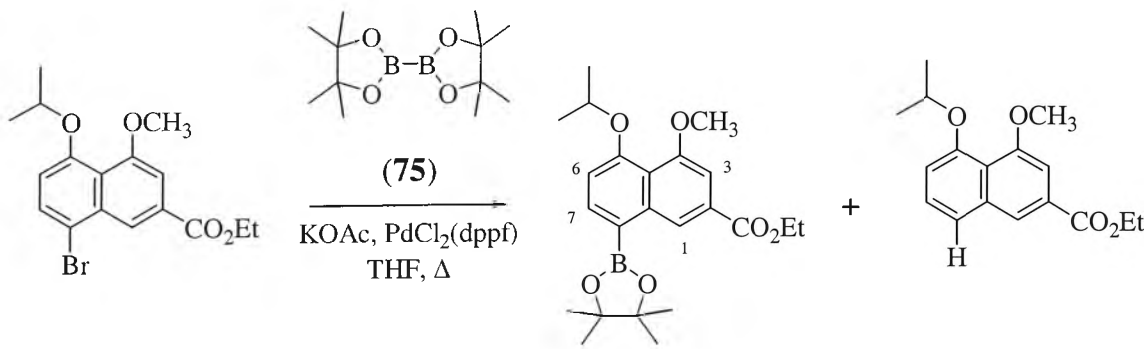


(73) (68) (65)

2.4.2.1 Preparation of boronic acids and esters

The boronate ester (**73**) was prepared in one step using bis(pinacolato)diboron (**75**),[102,103] (Scheme 2.16). In a typical reaction (**65**) was added to bis(pinacolato)diboron (**75**), dichloro[1,1'-bis(diphenylphosphino)ferrocene]palladium [$PdCl_2$(dppf)], and potassium acetate in THF, and heated at reflux for 24 hours. The crystalline boronate ester (**73**) was produced in 95 % yield, and only a short plug of silica was required to remove some baseline material for purification. TLC and mass spectral data revealed trace amounts of the debrominated starting material (**76**), with an *M*+1 peak at 289. The choice of KOAc as base was essential to the acceleration of the reaction; too weak a base would slow down the reaction, and too strong a base would induce the formation of biaryl by-products.[103,106]

Scheme 2.16



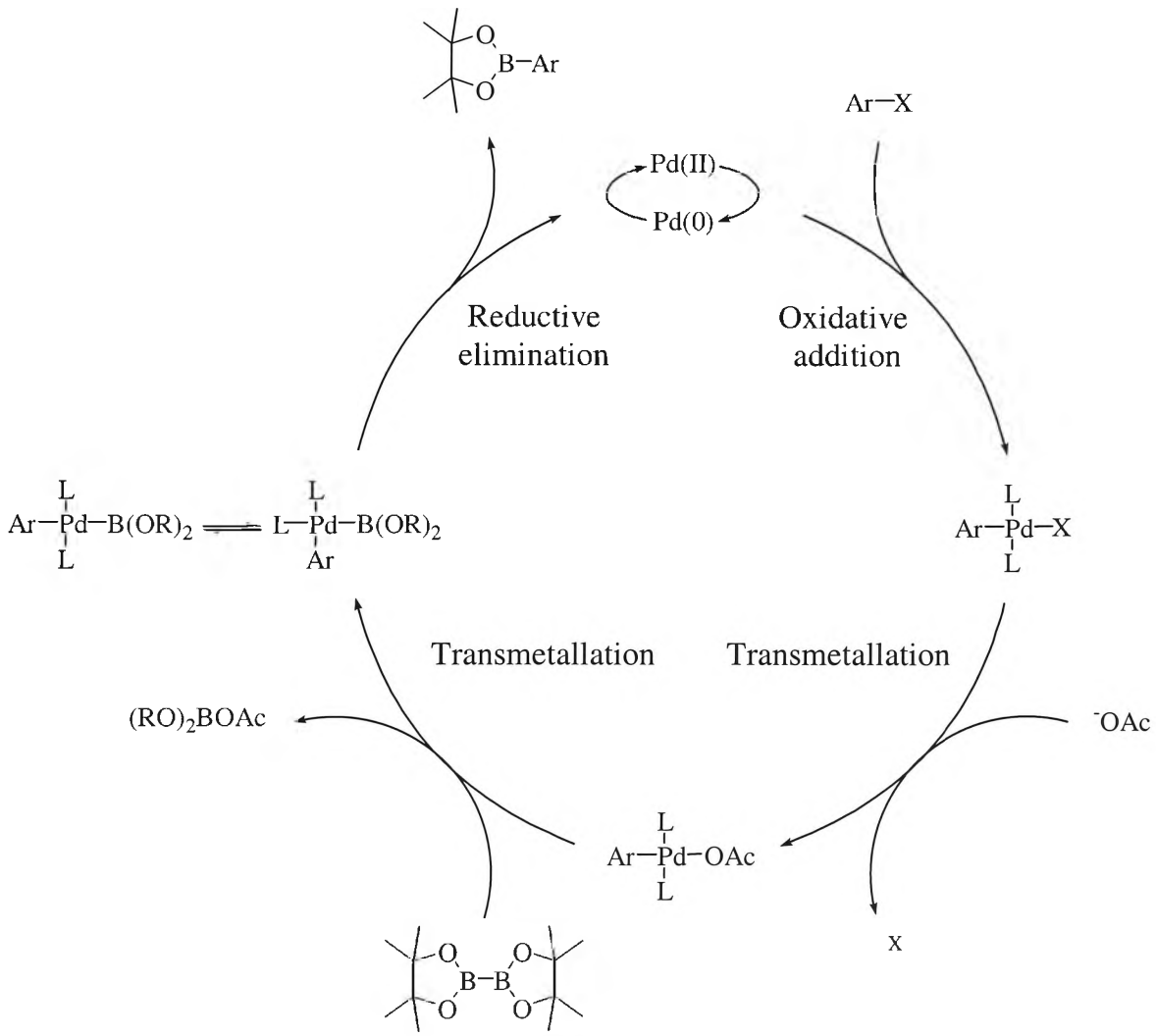
(65) (73) (95%) (76) trace

NMR data showed that there was restricted rotation about the C-B bond, as the symmetry of the boronate ester was broken. The alkyl groups of the boronate ester (**73**), that were expected to appear equivalent, actually appeared as two signals, seen in both

^{1}H and ^{13}C NMRs (δ 1.416 and 1.417, and δ 24.92 and 24.99). The aromatic hydrogens were expected to show an upfield shift with the replacement of the bromine with the less electronegative boron. Instead, the hydrogens in the 1, 6 and 7 positions shifted notably downfield to δ 9.19, δ 6.97, and δ 7.99 respectively. The hydrogen at position 3 showed a small upfield shift to δ 7.39 from δ 7.46 in (**65**). The 8-carbon showed another unexpected downfield shift, from δ 134.40 to δ 140.50, with the replacement of the bromine, as did carbons 7 and 8a. The 3-H signal in the ^{1}H NMR at δ 7.39 unexpectedly appeared as a singlet when a doublet, with coupling constant of 1.2 Hz, as seen for the doublet of 1-H, was expected. This was attributed to a resolution problem with the spectrometer used to obtain the NMR. Mass spectra showed a *M*+1 at 415 as the molecular ion, and a *m/z* at 373 (-42) was attributed to the loss of the isopropyl group.

Scheme 2.17 shows the proposed catalytic cycle of the cross coupling reaction of the pinacol ester of diboronic acid (**75**).[103]

Scheme 2.17

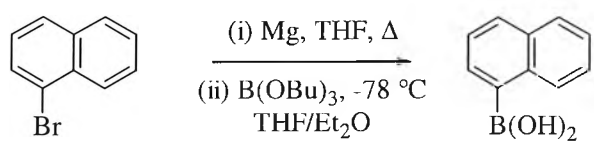


The oxidative addition and reductive elimination sequences are well understood,[97] unlike the transmetallation step, which is highly dependent on the organometallic species and conditions used.[86] The catalytic cycle begins with the oxidative addition of the brominated naphthyl (**65**), represented by Ar-X in Scheme 2.17, to a palladium(0) complex to yield an Ar-Pd(II)$(L)_2$-X adduct. Two transmetallation steps successively replace the halide group with the acetoxy group, and then the acetoxy group with the boronate ester. *Cis-trans* isomerisation can occur with the Ar-Pd(II)$(L)_2$-$B(OR)_2$ intermediate before reductive elimination of the naphthyl boronate ester (**71**), from the *cis* isomer, to regenerate the palladium(0) complex.[110]

The role of the base in this mechanism is not definitively understood; it is thought to accelerate the transmetallation rate by displacing the palladium halide complex to give an acetoxopalladium(II) species in solution, which is known to transmetallate with organoboron compounds in neutral conditions.[103] Another possibility is that the coordination of the negatively charged base to the boron atom increases its nucleophilicity for transmetallation to the palladium halide.[103] The high reactivity of the oxopalladium complex, formed by the base, toward transmetallation with organoboron compounds can explain the high yields obtained in this reaction. This high reactivity may be due to the reactivity of the Pd-O bond that consists of a soft acid and a hard base combination, and the high oxophilicity of the boron centre.[103]

Phenyl, and methyl boronic acid were purchased, while naphthyl boronic acid was prepared via a Grignard intermediate (Scheme 2.18).[98] Bromonaphthalene (**70**) was added to magnesium shavings, and a crystal of iodine in THF, and heated at reflux for 2 hours. The addition of tri-*n*-butylborate, followed by an acidic workup, and recrystallisation from water, resulted in 40% of (**77**).

Scheme 2.18



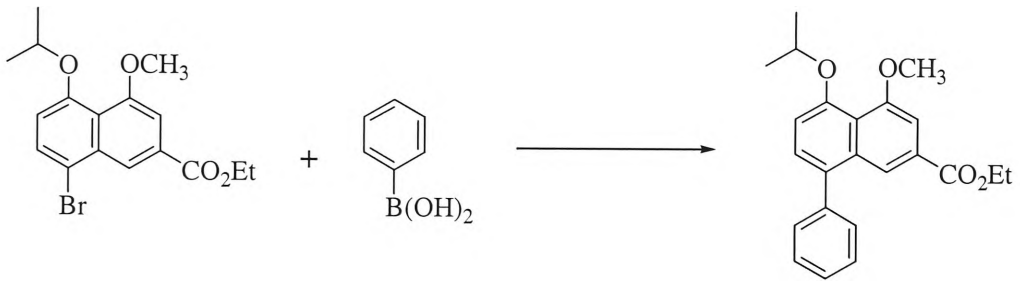
(**70**) (**77**)

2.4.2.2 Coupling of naphthyl halides and aryl boronic acids

In a typical Suzuki cross-coupling, the boronic acid was added to a mixture of the aryl halide, a palladium catalyst, and a base, and heated at reflux in a sealed tube or in standard reflux apparatus, under nitrogen for a minimum of 24 hours. The reaction was monitored by TLC until such a time it was determined that the reaction would not proceed further. This time scale of the reaction, on the whole, did not exceed 48 hours, but was dependent on reaction conditions, in particular the nature of the catalyst used.

The cross-coupling of phenyl boronic acid (**74**) of Scheme 2.19 revealed differences in reaction yields depending on the solvent and catalyst used as detailed in Table 2.3. Using $Ba(OH)_2$ and $Pd(PPh_3)_4$, the solvent in a series of phenyl boronic acid couplings with (**65**) was varied and the reaction yields monitored (entries 1-4). Compound (**65**) was spectroscopically identical to the product produced *via* the Stille cross coupling reaction (Section 2.4.1.2).

Scheme 2.19



(**65**) (**74**) (**68**)

Entry	Catalyst	Solvent	Temp. (°C)	Time (hrs)	Yield %
1	$Pd(PPh_3)_4$	DMF	160	48	78 %
2	$Pd(PPh_3)_4$	DME	90-100	48	69 %
3	$Pd(PPh_3)_4$	dioxane	110	32 - 36	85 %
4	$Pd(PPh_3)_4$	THF	90	32 - 36	85 %
5	$PdCl_2(dppf)$	THF	90	24	97 %

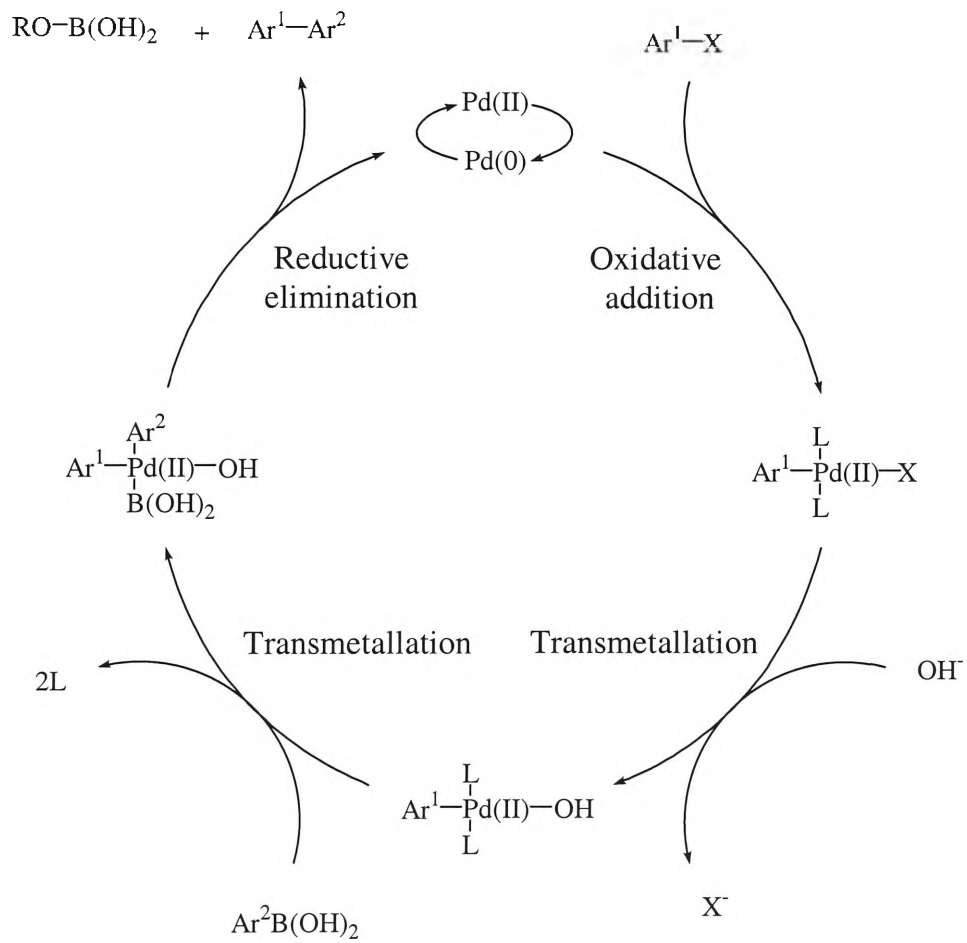
Table 2.3 Optimisation of the Suzuki cross-coupling of phenyl boronic acid (74) and the brominated naphthyl (65), Scheme 2.19. $Ba(OH)_2$ was used as the base in all reactions. Temperatures reported were oil bath temperatures.

Entries 1 and 2 show that DMF and DME produced reasonable yields of (**68**), but reaction times on average approached 48 hours. The use of DMF also required a much higher reflux temperature (oil bath 160 °C). The reaction appeared to halt after 48 hours, as monitored by TLC analysis, probably due to decomposition of the catalyst.

Tetrahydrofuran proved to be the best solvent for use (entry 4), and although dioxane had comparable yields with $Pd(PPh_3)_4$ as the catalyst (entry 3), the lower boiling point and accessibility of THF made it the preferable solvent to use. The variation of $PdCl_2$(dppf) as catalyst had the dual effects of increasing the reaction yield and reducing the reaction time.[101,103,111] Entry 5 of Table 2.3 produced the highest yield of product and involved only an extractive work-up, a short celite plug, and recrystallisation to purify (**68**). As a general rule, $Pd(PPh_3)_4$ was the preferred catalyst for general use as is was easily and economically synthesised,[112] but $PdCl_2$(dppf) was most frequently found to be the better catalyst in the cross coupling.

The mechanism for the Pd-catalysed Suzuki reaction of phenyl boronic acid (**74**) with the brominated naphthyl (**65**) is most likely similar to the cross coupling reaction of the pinacol ester of diboronic acid (**75**) (Section 2.4.2.1), involving an oxidative addition of Pd(0) to the aryl halide, transmetallation of the Ar^1-Pd(II)-X with the boronic acid and reductive elimination to give the cross-coupled product (Ar^1-Ar^2) as shown in Scheme 2.20.

Scheme 2.20

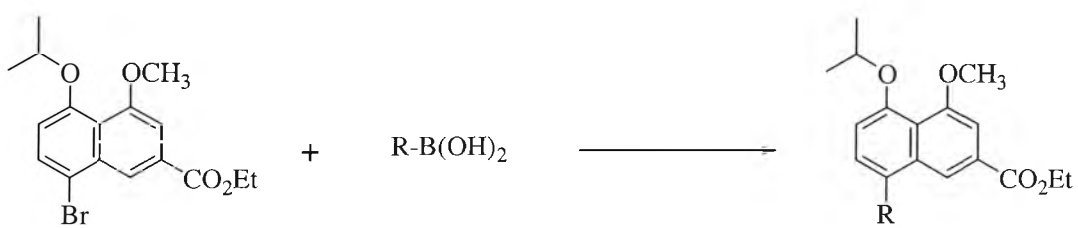


The base in the reaction is represented by OH^- in Scheme 2.20, and its strength is thought to play a role in the acceleration of the coupling rate.[96,97,103,106,113] The oxidative addition is often thought to be the rate determining step in the catalytic cycle, with the bromine of (**65**) being more reactive than a chlorine in this position, but less reactive than an iodine or triflate compound.[86]

Transmetallation between the Ar^1-$Pd^{(II)}(L)_2$-X adduct and the Ar^2-$B(OH)_2$ reagent does not occur readily due to the low nucleophilicity of the organic group on the boron atom.[86] However the nucleophilicity of the organic group on the boron atom can be enhanced by quaternization of the boron with negatively charged bases, such as $Ba(OH)_2$ in Scheme 2.19, allowing the transmetallation with the organopalladium(II) halide. The reductive elimination of the coupling partners (Ar^1 and Ar^2) from the Ar^1-$Pd^{(II)}(OR)B(OH)_2$-Ar^2 complex reproduces the palladium(0) complex. The Ar^1-$Pd^{(II)}(OR)B(OH)_2$-Ar^2 exists as *cis* and *trans* isomers, with the reaction taking place directly from the *cis* isomer, as shown in Scheme 2.20. The coupling of two aryl units is known to be more easily achieved than that of an aryl and alkyl, or two alkyl groups, suggesting participation of the π–orbital of the aryl group during the reductive elimination bond formation.[86]

The optimised conditions for the Suzuki cross-coupling of (**65**) and (**74**) in Table 2.3 were used as a basis for the synthesis of the naphthyl and methyl cross-coupled products (Scheme 2.21). $PdCl_2(dppf)$, $Ba(OH)_2$, and THF were initially used as catalyst, base and solvent respectively (Table 2.4).

Scheme 2.21



(**65**) R = naphthyl (**77**), methyl (**78**) → R = naphthyl (**72**), methyl (**79**)

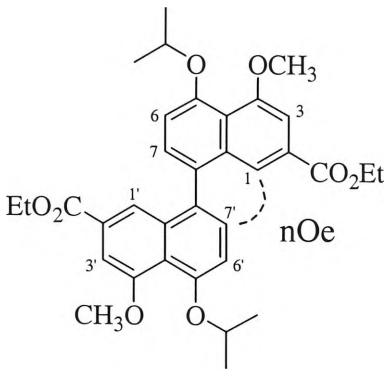
Entry	R	Catalyst	Base	Time (hrs)	Result
1	naphthyl	$PdCl_2(dppf)$	$Ba(OH)_2$	24	(**72**) 42 %
2	naphthyl	$Pd(PPh_3)_4$	$Ba(OH)_2$	48	(**72**) 87 %
3	naphthyl	$PdCl_2(dppf)$	KOAc	24	(**72**) 91 %
4	methyl	$PdCl_2(dppf)$	$Ba(OH)_2$	24	(**79**) 38 %
5	methyl	$PdCl_2(dppf)$	K_2CO_3	24	(**79**) 94 %

Table 2.4 Conditions used in the Suzuki cross-coupling of naphthyl and methyl boronic acid and the naphthyl core unit, Scheme 2.21. THF was used as the solvent in all reactions, and refluxed at an oil bath temperature of 90-100 °C.

The use of $PdCl_2(dppf)$ and $Ba(OH)_2$, entries 1 and 4, in the Suzuki cross-coupling yielded 42% and 38% of the naphthyl (**72**) and methyl (**79**) cross-coupled product, respectively. The substitution of $Pd(PPh_3)_4$ as the catalyst, in entry 2, more than doubled the yield of (**65**) to 87 %, however the reaction time increased to 48 hours. In entry 3, $PdCl_2(dppf)$ was retained as the catalyst, with KOAc as the base, increasing the product yield to 91%. Similarly replacing $Ba(OH)_2$ with K_2CO_3, in entry 5, increased the yield of the methyl cross-coupled product (**79**) to 94 %. The syntheses of (**72**) with K_2CO_3, and (**79**) with KOAc as the base were not attempted, though the results from such reactions may be interesting to determine the affect of the base on each substituted cross-coupled product.

The ^1H and ^{13}C NMR spectroscopy of (**72**) from Scheme 2.21 was compared to the product synthesised by the Stille cross-coupling and found to be identical. The spectroscopy of (**79**) showed an additional singlet at δ 2.65 and δ 19.75 in the ^1H and ^{13}C NMR alkyl regions respectively for the aryl methyl. The 1-H and 7-H showed shielding effects and signals downfield at δ 7.31, and δ 7.24 respectively to those seen in (**65**) at δ 8.56 and δ 7.70. The 3-H and 6-H were virtually unchanged at δ 7.43 and δ 6.94 respectively.

Further investigation of the low yielding reactions where $PdCl_2(dppf)$ and $Ba(OH)_2$ were used, entries 1 and 4 of Table 2.4, revealed the homo-coupled product (**80**) of the brominated naphthyl unit (**65**), depicted in Figure 2.4. The percentage of (**80**) was not accurately determined in these reactions, but as a qualitative gauge, TLC indicated that it was the major product compared to the desired cross-coupled product.



(**80**)

Figure 2.4 Structure of the homo-coupled product of the Suzuki cross-coupling reactions.

The structure of (**80**) was initially proposed after a *m/z* peak at 575 was identified by mass spectrometry, corresponding to *M*+1 for (**80**). The symmetry of the product was reflected in the ^{1}H NMR, which had the same number and types of signals as (**65**), but showed different proton chemical shifts. The 1-, 3- and 7-H of (**80**) were shifted upfield to δ 7.68, δ 7.38, and δ 7.37 respectively compared to δ 8.56, δ 7.46, and δ 7.70 in (**65**). A downfield shift from δ 6.87 to δ 7.13 was observed for 6-H. The dimeric nature of (**80**) was most convincingly demonstrated by an nOe detected between 1-H and 7'-H, as seen in Figure 2.4. Similarly to the naphthyl cross-coupled product (**72**), the methyl groups of the isopropyl appeared as two doublets, instead of one doublet as expected. The methylene hydrogens of the ethyl ester also appeared as two quartets instead of one quartet, however the methyl of the ethyl ester was unaffected, and appeared as a single triplet. The ^{13}C NMR showed two carbon signals for the methyl groups of the isopropyl, but no doubling of carbon signals for the methylene carbons of the ethyl ester. Another temperature series of ^{1}H NMRs, increasing in 10 °C increments from 20 °C to 50 °C, showed the doublets of the isopropyl methyl beginning to come together, but not totally collapsing into a single doublet. The proton signals of the methylene of the ethyl ester showed no change through the series of temperature regulated ^{1}H NMRs.

From the observations of the results of Table 2.4 it was speculated that the choice of base and catalyst may have had differing effects on not only the reaction rate and yield, but may also determine the nature of the products. In the coupling of (**65**) and (**74**), Scheme 2.19 and Table 2.3, $PdCl_2$(dppf) had already been demonstrated to be a more effective catalyst than $Pd(PPh_3)_4$ in increasing the rate of reaction and the product yield. However in the instance of the naphthyl cross-coupled product, the use of $PdCl_2$(dppf) produced a dramatically decreased yield of (**72**) (42%) compared to the same reaction where $Pd(PPh_3)_4$ was used (87%). The choice of base was also considered to play a role in the rate of the reaction, with stronger bases thought to accelerate the transmetallation rate of the coupling reaction.[97,103] Yet when the stronger base was used, $Ba(OH)_2$ instead of KOAc in the synthesis of (**72**) and K_2CO_3 in the synthesis of (**79**), the product

yields were poorer; 42% compared to 91% and 38% compared to 94% for (**72**) and (**79**) respectively.

The results appeared to show that the cross coupling reactions produced the optimal yield when a certain reaction rate was achieved with the combination of catalyst and base. Exceedingly strong catalysts and bases, which accelerate the rate of reaction beyond the optimal for cross coupling, could favour the oxidative addition step of Scheme 2.20, over the two successive transmetallation steps necessary before the reductive elimination. The dominance of the oxidative addition would lead to the unwanted by-product, such as (**80**), rather than the cross-coupled product. Weaker catalysts and bases, could retard the rate of reaction, and although this would minimise the extent of (**80**) in the reaction, it could also see the appearance of debrominated starting material (**76**). This was similarly seen in the cross coupling reaction of the pinacol ester of diboronic acid of Scheme 2.16.

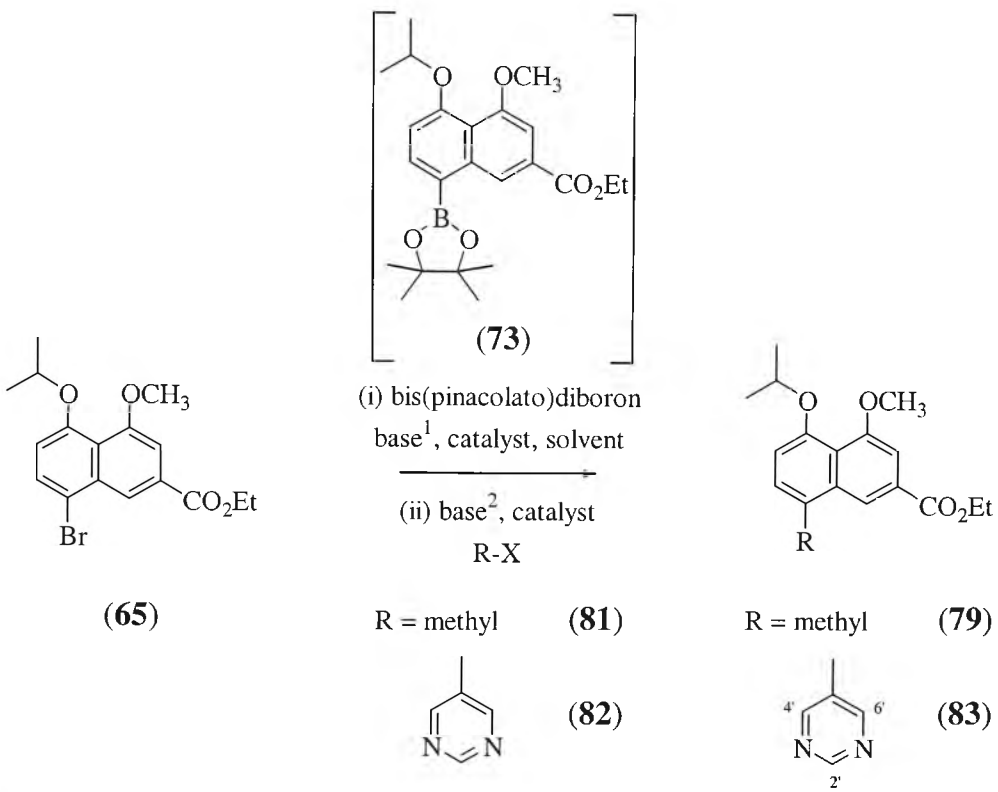
2.4.2.3 Coupling of naphthyl boronate esters and aryl halides

The Suzuki reactions for the methyl, pyrimidine, and pyridine cross-coupled products were carried out with the naphthyl boronate ester (**73**) and iodomethane, 5-bromopyrimidine, and 2-bromopyridine respectively. These substituents were most easily available as organic halides, consequently the synthesis of their boronic acids or esters was not pursued.

The cross coupling reactions of Scheme 2.22 were carried out as one-pot *in situ* reactions,[101] without isolation and purification of the boronate ester (**73**). The formation of (**73**) was monitored by TLC and mass spectra before the addition of the aryl or alkyl halide substituent. In a typical reaction the brominated naphthyl (**65**) was added to bis(pinacolato)diboron, base (1), and catalyst in THF, and heated at reflux, under nitrogen, for 24 hours. Base (2), iodomethane or an aryl bromide, and a second equivalent of catalyst were added and heated at reflux, under nitrogen, for a further 24 hours. TLC monitoring of the reaction after the initial 24 hours showed complete transformation of the brominated naphthyl (**65**) to the boronate ester (**73**), with occasional trace evidence of the corresponding debrominated starting material (**76**). A mild base is usually used for the first step of the reaction to minimise the homo-coupled by-product (**76**),[101-103] before a stronger base is used to both hydrolyse the naphthyl boronate ester to the corresponding acid for reaction, and to participate in the cross-

coupling catalytic cycle of Scheme 2.20. Strangely, the use of a strong base in the second stage of the reaction did not appear to have an effect in favouring the unwanted homo-coupled by-product (**76**).

Scheme 2.22



Entry	R	Base[1]	Catalyst	Base[2]	Result
1	methyl	KOAc	$Pd(PPh_3)_4$	$Ba(OH)_2$	(**79**) 95 %
2	methyl	KOAc	$PdCl_2(dppf)$	$Ba(OH)_2$	(**79**) 80 %
3	methyl	K_2CO_3	$PdCl_2(dppf)$	$Ba(OH)_2$	(**80**) 92 %
4	pyrimidine	K_2CO_3	$PdCl_2(dppf)$	$Ba(OH)_2$	(**83**) 40 %
5	pyrimidine	KOAc	$PdCl_2(dppf)$	$Ba(OH)_2$	(**83**) 77 %

Table 2.5 Conditions used in the one-pot Suzuki cross-coupling of (65) with iodomethane and 5-bromopyrimidine, Scheme 2.22. THF was the solvent in all the reactions.

Although a high yielding synthesis of the methyl cross-coupled product (**79**) had already been achieved, the differences in yield between the two basic strategies of Scheme 2.15 were calculated to determine which was the better unit to have as the boronic acid. The comparison of the 94 % yield from the cross coupling of the brominated naphthyl (**65**) with methyl boronic acid (**78**), in Scheme 2.21 and Table 2.4, against either of the yields from the iodomethane reactions of Table 2.5, entries 1 and 2, do not definitively differentiate between which coupling partner is the better moiety to have as the boronic

acid. Although different bases were selected because of the one-pot *in situ* requirements, the reaction where $Pd(PPh_3)_4$ was used produced a virtually identical yield of the cross coupled product (95 %). The slightly lower yield of 80 % when $PdCl_2$(dppf) was used, the same catalyst used in the methyl boronic acid coupling, may be attributable to by-products, such as the homo-coupled naphthyl (**80**), although this was not confirmed by chromatography and spectroscopy. When K_2CO_3 was used as the first base with $PdCl_2$(dppf), in entry 3, 92% of the homo-coupled naphthyl (**80**) was isolated, without any trace of the cross-coupled methyl product(**79**).

In the cross coupling of 5-bromopyrimidine, the use of K_2CO_3, as the first base, resulted in a 40 % yield of the cross-coupled product (**83**), and a 48 % yield of the homo-coupled product (**80**). Separation of the reaction mixture by silica chromatography, followed by NMR spectroscopy and mass spectrometry, confirmed the yields and structures of (**83**) and (**80**). The use of KOAc in place of K_2CO_3 in the pyrimidine cross-coupling increased the yield of cross coupled product to 77%; the yield of the unwanted by-product was not determined in these later reactions.

The mass spectra of (**83**) showed a *M*+1 at 367, and a *m/z* at 325, corresponding to the loss of the isopropyl group. The symmetry afforded by the use of the 5-pyrimidine substituent made for easy identification in ^{1}H NMR, as seen in Figure 2.5. The expected shielding effects of the pyrimidyl ring were seen in the upfield shift of 1-H and 7-H from δ 8.56 to 7.96 and δ 7.70 to 7.32 respectively, compared to (**65**). The shifts compared to the boronate ester (**73**) are also illustrated in Figure 2.5. 3-H showed negligible difference in signal between the pyrimidine cross-coupled product (**83**) and the brominated naphthyl (**65**), while a downfield shift was observed in 6-H from δ 6.87 to 7.07. The 4'- and 6'-H appeared as a singlet of two hydrogens at δ 8.83 and the 2'-H appeared at δ 9.27 as expected for the pyrimidyl aryl substituent. HSQC and HMBC correlated the identified protons of the pyrimidine ring to one carbon signal at δ 157.41, that represented three carbons of the pyrimidine ring; the *ipso* carbon was at δ 134.34. NOESY experiments also confirmed an nOe between 7-H of the naphthyl ring system and the 4'- or 6'-H of the pyrimidine.

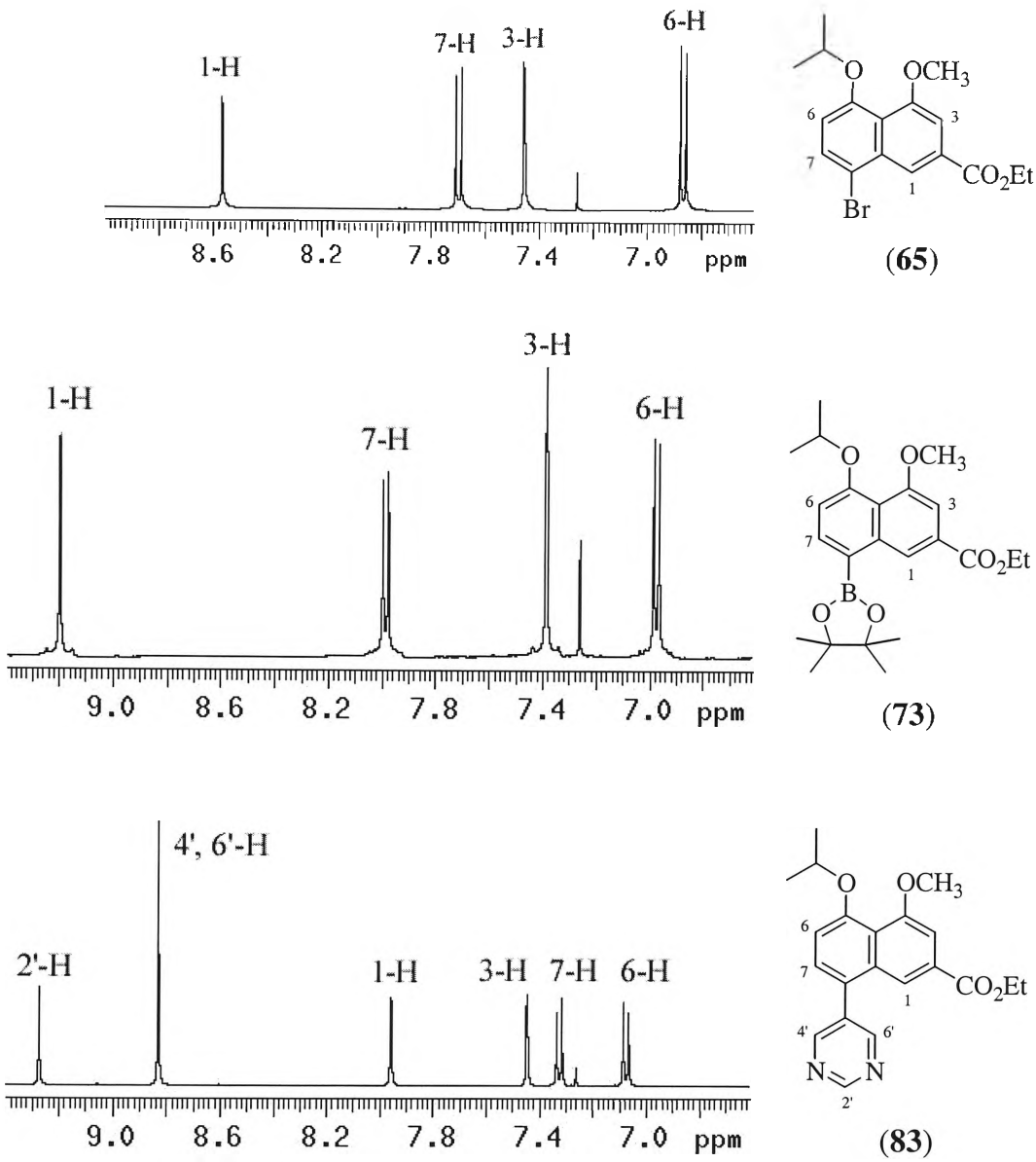


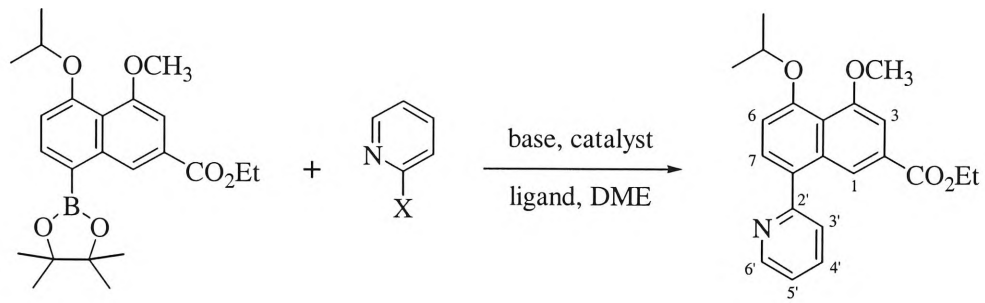
Figure 2.5 The ^{1}H NMRs of the brominated naphthyl (65), the boronate ester (73) and the pyrimidine cross-coupled product (83) illustrate the shifts in the aromatic region.

In the synthesis of the pyridine cross-coupled product, the optimised conditions of entry 5 of Table 2.4, using $PdCl_2(dppf)$, and KOAc and $Ba(OH)_2$ as the first and second base, respectively, in THF, failed to yield any cross-coupled product, and only the intermediate boronate ester (**73**) was recovered. Changing the solvent and the second base to DME and K_2CO_3, respectively, resulted in no change to the reaction outcome (entry 1, Table 2.6). An increase in the time of heating at reflux after the addition of 2-bromopyridine, the second base, and fresh catalyst, from 24 to 48 hours showed no formation of the pyridine cross coupled product. However, TLC and mass spectral data

of these longer reactions began to show the appearance of (**76**), the debrominated, or in this case deboronated form of (**65**) and (**73**) respectively.

The reaction was modified by isolating and purifying the naphthyl boronate ester (**73**), before proceeding to the reactions of Scheme 2.23. The substitution of $Pd(OAc)_2$ and PPh_3 as catalyst and coordinating ligand, instead of $PdCl_2(dppf)$, and DME as solvent instead of THF, as suggested from literature Suzuki couplings of chloropyridines,[96,102] proved to be successful, if not high yielding as seen in entries 4 and 5 of Table 2.6. As expected, the coupling of 2-bromopyridine resulted in a greater yield than that of 2-chloropyridine.

Scheme 2.23



(**73**) (**84**)

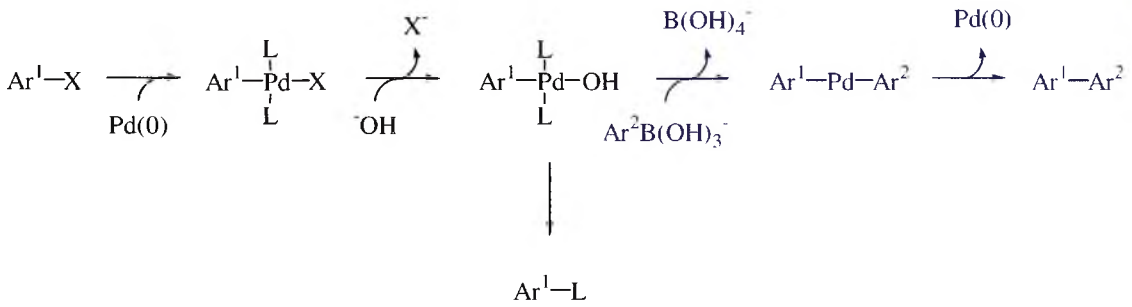
Entry	Base	X	Catalyst and Ligand	Result
1	K_2CO_3	Br	$PdCl_2(dppf)$	(**73**)
2	KOAc	Br	$Pd(OAc)_2$	(**73**)
3	$Ba(OH)_2$	Br	$Pd(OAc)_2$	(**73**)
4	K_2CO_3	Cl	$Pd(OAc)_2$, PPh_3	(**84**) 36 %
5	K_2CO_3	Br	$Pd(OAc)_2$, PPh_3	(**84**) 45 %

Table 2.6 Conditions used in the Suzuki cross-coupling of the naphthyl boronate ester and 2-halopyridine, Scheme 2.23. DME was used as the solvent in all reactions, and refluxed at 90-100 °C, under nitrogen.

The TLC analyses of entries 4 and 5 of Table 2.6 revealed the formation of a secondary product that had a retention time similar to that of the phenyl cross-coupled product (**68**). Mass spectra showed a *M*+1 peak at 366 for (**84**) and a *m/z* of 365, that corresponded to a *M*+1 for the phenyl cross-coupled product (**68**). High resolution gas chromatography and mass spectra confirmed the elemental composition of the *m/z* 365 and 366 peaks as $C_{22}H_{24}O_4$ (**68**) and $C_{22}H_{23}NO_4$ (**84**), respectively. The yield for the phenyl cross-coupled product in these reactions was not calculated, but TLC analysis indicated that it was approximately equivalent to that of the pyridine cross-coupled

product (**84**). Presumably the phenyl cross-coupled product (**68**) is produced when the reductive elimination occurs after the first transmetallation in Scheme 2.20. The phenyl ligand, L, from the triphenylphosphine of the catalyst is coupled with Ar^1 from (**73**) to form the phenyl cross-coupled product, instead of the Ar^1-Ar^2 pyridine cross-coupled product, shown in grey in Scheme 2.24. This 'product scrambling' has been previously noted in Suzuki,[104,105] and Stille reactions,[114] and may be minimised by reducing the catalyst quantity,[105] or using phosphine free catalytic systems.[106]

Scheme 2.24



No explanation could be given for the failure of $PdCl_2$(dppf) to catalyse the cross coupling of the (**73**) and 2-bromopyridine. It can be speculates that the positioning of the nitrogen *ortho* to the halogen may have involved it in some coordination complex with the catalyst $PdCl_2$(dppf) that was unfavourable for coupling. Repeating the reaction with $Pd(PPh_3)_4$ may reveal if indeed it was the bis(diphenylphosphino)ferrocene ligand that was the cause of the problem.

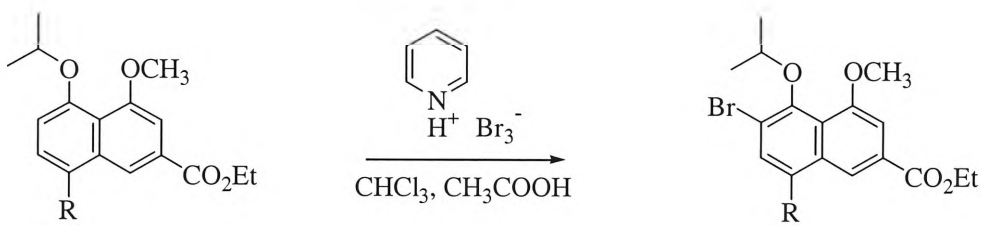
In addition to the previously mentioned mass spectral data, the signature splitting patterns of the pyridine aryl unit helped to establish the structure of (**84**). The 5'-H appeared as a doublet of doublets of doublets, with 7.4 Hz *ortho* and 0.8 Hz *meta* coupling to the 4' and 6'-H, and 3'-H, respectively. The resolution was sufficient to see individual peaks, but possibly insufficient for accurate detection of the small *meta* coupling which was anticipated to be greater than 0.8 Hz. The splitting pattern of 3'-H was obscured by the peaks of the doublet of 7-H, and only resolved as a doublet of 7.6 Hz, coupled to 4'-H. Another doublet of doublets of doublets, which visually appeared as a doublet of triplets characteristic of pyridines, correlated to 4'-H, with 7.6 Hz *ortho* and 1.6 Hz *meta* coupling to 3' and 5'-H, and 6'-H, respectively. The 6'-H was not sufficiently resolved, and appeared as a broad doublet, with approximately 4.0 Hz coupling.

2.5 *Ortho*-directed bromination

Ortho bromination at the 6 position of the cross-coupled products was proposed to enable the dimeric series of targeted compounds to be synthesised from the monomeric cross-coupled series (Scheme 2.1). This bromine handle would allow access to the Stille and Suzuki homo-coupling reactions to form the binaphthyl moiety of the targeted dimeric compounds.

In a typical reaction, a suspension of pyridinium tribromide in glacial acetic acid was added dropwise to a stirred solution of the R-substituted naphthyl product, in chloroform at 0 °C (Scheme 2.25).[65] In the case of (**83**) and (**84**), the chloroform was pre-acidified to compensate for the basic nature of the pyrimidine and pyridine substituents.

Scheme 2.25



R = phenyl (**68**)
naphthyl (**72**)
methyl (**79**)
pyrimidine (**83**)
pyridine (**84**)

R = phenyl (**85**) (76%)
naphthyl (**86**) (97%)
methyl (**87**) (91%)
pyrimidine (**88**) (98%)
pyridine (**89**) (81%)

R	1-H		3-H		7-H	
	monomer	*ortho*-Br	monomer	*ortho*-Br	monomer	*ortho*-Br
phenyl	8.22	8.17	7.48	7.43	7.36	7.65
naphthyl	7.74	7.70	7.39	7.43	7.41	7.47
methyl	7.31	8.26	7.43	7.46	7.24	7.54
pyrimidine	7.96	7.93	7.44	7.52	7.32	7.64
pyridine	8.39	8.31	7.42	7.46	7.54	7.80

Table 2.7 ^{1}H NMR shifts (δ ppm) for the 1-H, 3-H and 7-H of the R-substituted cross-coupled products, and their *ortho*-brominated derivatives.

The yields of all the *ortho*-directed brominations were in excess of 75 %, with that of the pyrimidine cross-coupled monomer (**88**) being almost quantitative. The products were purified by column chromatography and recrystallised from isopropanol.

The most evident indication of bromine incorporation in ^{1}H NMR was expected to be the disappearance of the naphthyl 6-H, and the change of the 7-H from a doublet with *ortho* coupling, to a singlet. In (**85**) the 7-H appeared as a singlet, shifted downfield, at δ 7.65 from the doublet of 8 Hz at δ 7.36 in (**68**). The 1-H and 3-H showed small upfield shifts to δ 8.17 and δ 7.43 compared to δ 8.22 and δ 7.48 in (**68**), respectively. Table 2.7 compares the 1-H, 3-H and 7-H proton shifts between the R-substituted cross-coupled products and the corresponding *ortho* brominated derivatives. All of the brominated products with aromatic substitutions showed upfield shifts in the 1-H, and downfield shifting in the 7-H. The small shifting of the 3-H was downfield in all of the R-substitutions of Table 2.7, except for the previously mentioned phenyl cross-coupled product.

Figure 2.6 shows the significant downfield shifting by 0.95 ppm of the 1-H of (**87**). This was in contrast to the products with aromatic substituents which showed upfield shifts for 1-H.

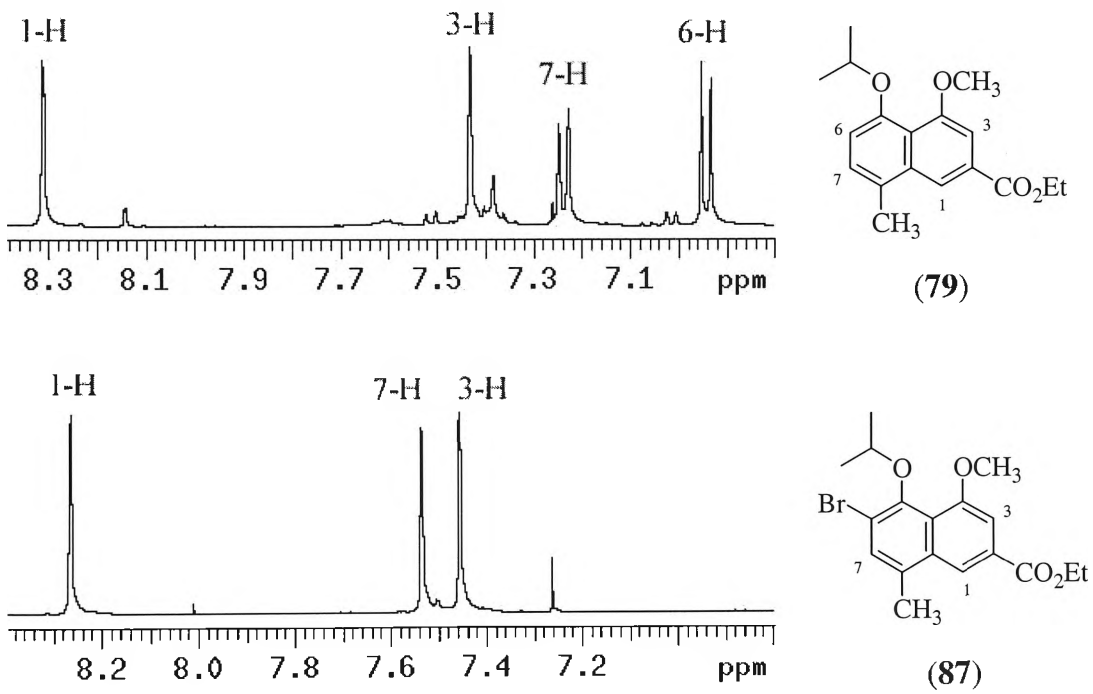


Figure 2.6 ^{1}H NMRs of the methyl cross-coupled product (79), and its *ortho*-brominated derivative (87) illustrating the shifts in the aromatic region.

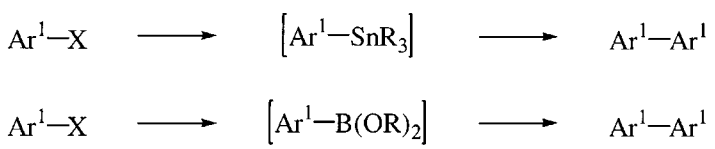
The coupling constants between 1-H and 3-H in the methyl substituted products also showed a notable change from 0.8 Hz in (**79**) to 1.6 Hz in (**87**). These coupling constants remained fairly consistent for the compounds with aromatic substituents

before and after the *ortho*-directed bromination. Figure 2.6 also shows the loss of the 6-H and the change of the 7-H from a doublet to a singlet.

2.6 Dimerisation reactions

Dimerisation, or homo-coupling, reactions are essentially the same as cross-coupling reactions; and in most cases can be achieved as a one-pot reaction where the stannylated or boronate intermediate does not necessarily need to be isolated (Scheme 2.26). A dimerisation can have different demands to a cross-coupling reaction; in a hetero-coupling reaction the choice of catalyst and base can play a major role in the yield of the product and the minimisation of dimerised products. However, this consideration is unnecessary in homo-coupling reaction, and bases and catalysts are often actually chosen to accelerate this process.

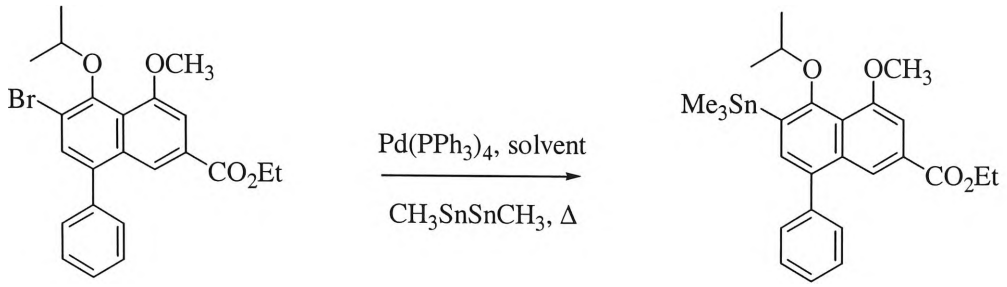
Scheme 2.26



2.6.1 Stille homo-couplings

Several attempts were made to dimerise the phenyl substituted monomer (**68**) using the Stille coupling reaction (Scheme 2.27). In a typical reaction, the *ortho*-brominated derivative was added to $Pd(PPh_3)_4$ and bis(trimethyltin), in an organic solvent and heated at reflux, under nitrogen, for 24-48 hours.[68,115] The use of toluene and THF as the organic solvent produced neither the dimerised product, nor the stannylated intermediate (**90**), and only approximately 50 % of the starting material (**85**) was recovered, with some polymeric material. The substitution of dioxane as the solvent in the reaction saw the appearance of *m/z* peaks at 527 and 529, corresponding to *M*+1 peaks (^{118}Sn and ^{120}Sn respectively) for (**90**), but no anticipated peak at *m/z* 727 for the dimer product. The yield of this reaction was approximately 40 %. With the concurrent success of the Suzuki reaction at this time, and the low initial yield in this Stille reaction, the decision was made to conclude the Stille homo-coupling reactions.

Scheme 2.27

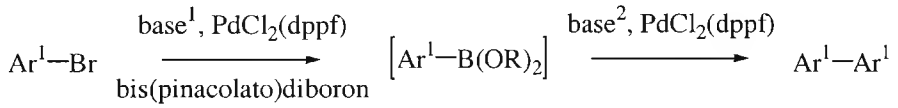


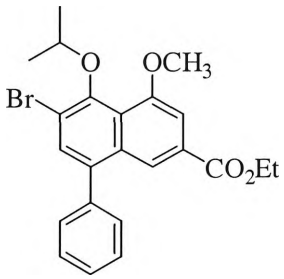
(85) **(90)**

2.6.2 Suzuki homo-couplings

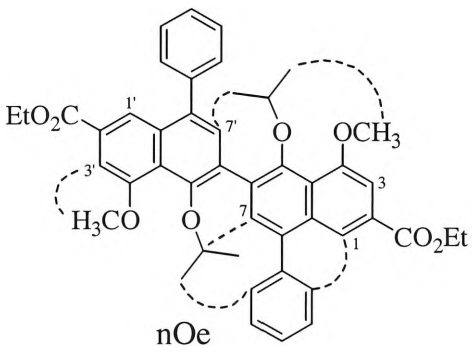
The Suzuki dimerisations were conducted as one-pot reactions where the boronate ester was not isolated and purified before the addition of the second aryl halide. In the homo-coupling of the phenyl substituted monomer (**74**), *via* its brominated derivative (**85**), different bases were used to evaluate their effect on the yield of the dimerised product (**91**) (Scheme 2.28, Table 2.8).

Scheme 2.28





(85)



(91)

Initially CsF was used as both the first and second base in the homo-coupling[116] (entry 1), yielding 15 % of (**91**). This yield was improved by the change to KOAc as the first base and $Ba(OH)_2$ as the second base. The use of these bases is similar to the strategy of the Suzuki cross-coupling (Section 2.4.2.3), where the first base was chosen as a relatively weak base to ensure the formation of only the boronate ester and to minimise

the formation of the unwanted homo-coupled product, before the second, stronger, base was chosen to accelerate the cross-coupling reaction

Entry	Base[1]	Base[2]	Catalyst	Ar^1-Ar^1
1	CsF	CsF	$PdCl_2$(dppf)	15 %
2	KOAc	$Ba(OH)_2$	$PdCl_2$(dppf)	61 %
3	$Ba(OH)_2$	$Ba(OH)_2$	$PdCl_2$(dppf)	64 %
4	K_2CO_3	K_2CO_3	$PdCl_2$(dppf)	50 %

Table 2.8 Conditions used in the one-pot Suzuki dimerisation of the *ortho*-brominated phenyl substituted monomer (85), Scheme 2.28. THF was the solvent in all the reactions.

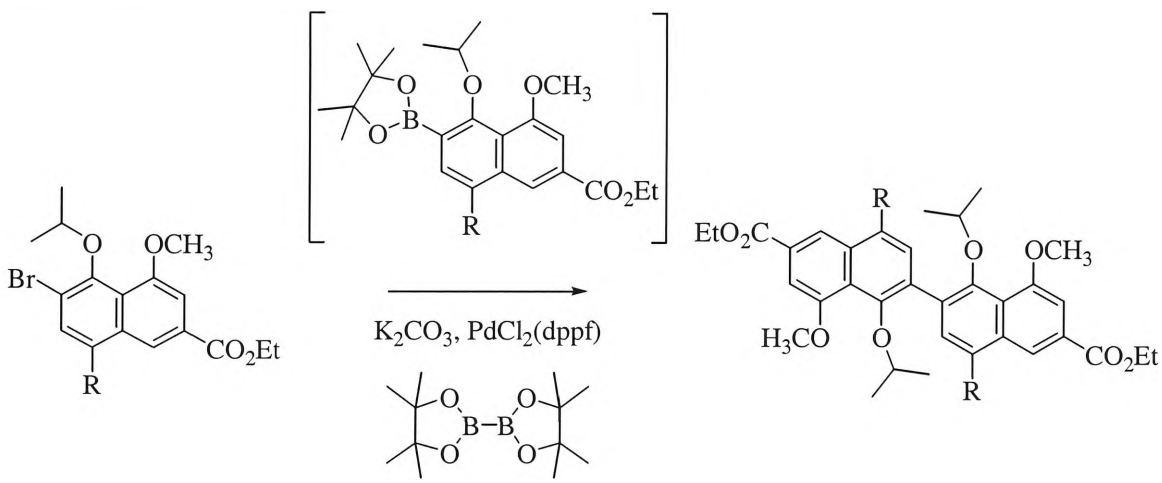
In entry 2 of Table 2.8, the use of KOAc and $Ba(OH)_2$ resulted in a 61 % yield of (**91**); the percentages of the boronate ester and the dimer product before the addition of the second equivalent of the *ortho*-brominated phenyl substituted monomer (**85**) were not determined. This yield was further optimised to 64 % with the sole use of $Ba(OH)_2$ in the coupling reaction. Despite this optimisation, the use of $Ba(OH)_2$, as the sole base or as the second base, proved to be inconsistent and unpredictable. While in most reactions, the dimer product was isolated without difficulty, several reactions resulted in the recovery of only a small amount of the phenyl cross-coupled product, polymeric material, and an uncharacterised product that was more polar than the expected dimer (**91**). The use of K_2CO_3, in entry 4, yielded only 50 % of (**91**), but proved to be more consistent in the synthesis of the dimer product.

The mass spectra of (**91**) showed a *m/z* at 727, corresponding to the *M*+1 of the dimer product, and *m/z* peaks at 685 and 643, which were assigned as the loss of one and both isopropyl groups, respectively. The ^{1}H NMR showed a notable upfield shifting in the proton signals of the isopropyl hydrogens; the methyl doublet of the isopropyl group was shifted from δ 1.38 in (**85**) to δ 1.11 in the dimer, and the methine quintet was shifted from δ 4.56 to δ 4.11 in (**91**). The naphthyl hydrogens (1, 1', 3, 3', 7, 7') all show downfield shifts; the most pronounced of these are those of the 7 and 7' hydrogens which are shifted from δ 7.65 in (**85**) to δ 7.96 in (**91**). These shifts are most likely due to the deshielding effects of the opposite naphthyl ring systems. The 3 and 3' hydrogens appeared as a singlet rather than the expected doublet, with 1.6 Hz coupling, that the 1 and 1' hydrogens displayed. The dimeric nature of (**91**) was further supported by NOESY experiments that detected nOes between the methyl groups of the isopropyl groups and both the 7 and 7'-H, and the phenyl substituents (Scheme 2.28). A nOe was

also detected between the isopropyl methine and the 7-H. These nOes would not be possible in the monomeric product (**68**), and are detected between the isopropyl of one naphthyl system and the phenyl substituent, and naphthyl aromatic hydrogens of the second naphthyl system, as illustrated by the dotted curves on (**91**) in Scheme 2.28.

The conditions of the synthesis of the phenyl dimer product (**91**) in entry 4 of Table 2.8, with K_2CO_3 as the base, $PdCl_2$(dppf) as the catalyst, and THF as the solvent in the Suzuki homo-coupling, were used in the synthesis of the dimeric forms of the naphthyl, pyrimidyl and methyl cross-coupled products (Scheme 2.29). In a typical reaction, an *ortho*-brominated substituted monomer (Ar^1-Br) was added to bis(pinacolato)diboron, K_2CO_3, and $PdCl_2$(dppf) in THF, and heated at reflux, under nitrogen, for 24 hours, after which a second equivalent of the *ortho*-brominated monomer, K_2CO_3 and $PdCl_2$(dppf) were added and heated at reflux for a further 24 hours. After the initial 24 hours, TLC analysis indicated the formation of both the boronate ester and the dimer product.

Scheme 2.29



R = phenyl	(**85**)	R = phenyl	(**92**)	R = phenyl	(**91**)
naphthyl	(**86**)	naphthyl	(**93**)	naphthyl	(**96**)
pyrimidine	(**88**)	pyrimidine	(**94**)	pyrimidine	(**97**)
methyl	(**87**)	methyl	(**95**)	methyl	(**98**)

The yields of the dimerised products, and other side products for the homo-coupling at the completion of the reaction, are presented in Table 2.9. Extending the reaction time after the addition of the second equivalent of R-substituted monomer for the phenyl, naphthyl and methyl products to 48 hours, had no effect on the yield of dimerised product; the yields of the boronate esters and monomeric products were not determined.

A longer reaction time was not attempted for the synthesis of (**97**). The homo-coupling of the *ortho*-brominated pyridine derivative (**89**) was not completed at this time.

R	Yield (dimer)	Yield (boronate)	Yield (monomer)
phenyl	(**91**) 50 %	(**92**) 9 %	(**68**) 38 %
naphthyl	(**96**) 51 %	(**93**) 18 %	(**72**) 24 %
pyrimidine	(**97**) 71 %	(**94**) 6 %	(**83**) 18 %
methyl	(**98**) 51 %	(**95**) 14 %	(**79**) 23 %

Table 2.9 Yields of the dimerised, boronate ester, and R-substituted monomeric naphthyl (debrominated) products after the Suzuki homo-coupling, Scheme 2.29.

It was uncertain why longer reaction times did not result in the formation of further dimer product if, at the end of 24 hours after the addition of the second equivalent of the *ortho*-brominated derivatives, there was still boronate ester that could participate in the coupling reaction. TLC analysis and mass spectra indicated that at the completion of the reaction, 24 hours after the addition of the second equivalent of all reagents except bis(pinacolato)diboron, there was no evidence of any *ortho*-brominated monomer. These reactions were carried out with two equivalents of the bis(pinacolato)diboron reagent (**75**). In a similar synthesis of a dimeric pyranonaphthoquinone, Brimble *et al.* used four equivalents of (**75**), and CsF as a base, and reported a 53 % yield of the dimer product, and a 19 % yield recovery of the boronate ester that could be recycled.[116] However, in this procedure a second equivalent of the brominated starting material was not added. These results seemed to suggest that the naphthyl bromide moiety (Ar^1-Br) was more reactive than the naphthyl boronate ester [Ar^1-$B(OR)_2$], oxidatively entering the palladium catalytic cycle of Scheme 2.20 more readily than the boronate moiety. It is likely that the dimerisation reaction was not actually a true homo-coupling, but rather a hetero-coupling of Ar^1-Br and Ar^1-$B(OR)_2$. As the boronate ester was formed in the first step of the reaction (Scheme 2.28), it presumably coupled with some of the brominated naphthyl that had entered the catalytic cycle as Ar^1-Pd(II)$(L)_2$-X, but not reacted with the base and bis(pinacolato)diboron (Scheme 2.17).

We hypothesise that the dimer synthesis could be conducted in a one-step reaction where only half an equivalent of (**75**) would be used, thus half of the brominated starting material would be transformed to the boronate ester, to react with the remaining half equivalent of brominated starting material. The reaction of (**65**) with bis(pinacolato)diboron to yield 95 % of (**73**) (Scheme 2.16) was extremely efficient, and

if a similar rate and yield of reaction could be expected for the *ortho*-brominated derivatives of Scheme 2.29, this may prove to be a more efficient route to the synthesis of the dimeric compounds, ensuring the total reaction of the boronate moieties. The choice of base may also need further consideration, as it may play a role in accelerating the rate of reaction. A faster rate of reaction could reduce the yields of the monomeric forms of the products found in the reactions, which are likely to come from some aryl-palladium complex that does not complete the reductive elimination step of the catalytic cycle (Scheme 2.20) to form a coupled product.

The mass spectra of the methyl dimerised product (**98**) showed an *M*+1 peak at 603 and *m/z* peaks at 561 and 519, corresponding to the loss of one and both isopropyl groups, respectively. The naphthyl (**96**) and pyrimidyl dimer products (**97**) only showed *M*+1 peaks at 827 and 731, respectively, and did not have peaks corresponding to fragments with the loss of one or both isopropyl groups. The ^{1}H NMRs showed similar patterns in the aromatic naphthyl hydrogen shifts (Table 2.10).

R	1-H		3-H		7-H	
	ortho-Br	dimer	*ortho*-Br	dimer	*ortho*-Br	dimer
phenyl	8.17	8.22	7.43	7.43 *	7.65	7.96
naphthyl	7.70	7.74 *	7.43	7.40	7.47	8.04
methyl	8.26	8.34	7.46	7.48 *	7.54	7.84
pyrimidine	7.93	7.97 *	7.52	7.53 *	7.64	7.97

Table 2.10 ^{1}H NMR shifts (δ ppm) for the 1-H, 3-H and 7-H of the *ortho*-brominated derivatives and the corresponding dimeric derivatives. * denotes proton signals that did not resolve as doublets, and appeared as singlets.

The shifts of the 1 and 1' hydrogens were consistently downfield, however small, in the dimerised products, compared to the *ortho*-brominated derivatives. The 3 and 3' hydrogens were effectively the same. The 7 and 7' hydrogens showed the greatest downfield shifting in the dimer compounds due to their position *ortho* to the biaryl axis, and the stronger deshielding effects of the opposite naphthyl ring system. The 1, 1', 7 and 7' hydrogens appeared as one peak of 4H integration at δ 7.97 in the ^{1}H NMR of the pyrimidine dimer product (**97**), and were identified by HQSC showing correlation to two separate carbons of the ^{13}C NMR.

The isopropyl hydrogens of the naphthyl, methyl and pyrimidine dimer products showed the same strong upfield shifting as previously described in the phenyl dimer product (**91**). The largest shifts were seen in the isopropyl methine hydrogens (Table

2.11), on average δ 0.46, while the average shift of the isopropyl methyl hydrogens was δ 0.28. This effect is most likely due to the close proximity of the isopropyl methine to the methoxy in the 4 and 4' position. The isopropyl methyl hydrogens may also be affected to a smaller degree, or it could be likely that their rotation brings them close enough to be electronically affected by the oxygen of the methoxy group.

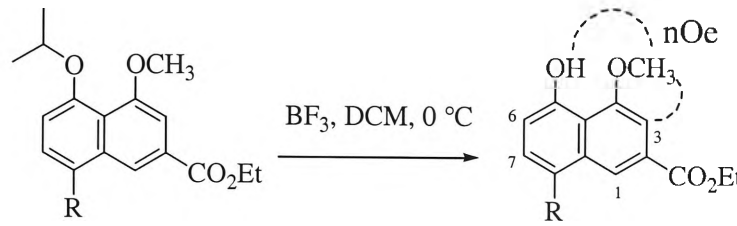
R	isopropyl methyl		isopropyl methine	
	ortho-Br	dimer	*ortho*-Br	dimer
phenyl	1.38	1.11	4.56	4.11
naphthyl	1.43	1.13	4.64	4.24
methyl	1.32	1.03	4.47	3.93
pyrimidine	1.39	1.12	4.58	4.15

Table 2.11 ^{1}H NMR shifts (δ ppm) for the isopropyl hydrogens of the ortho-brominated derivatives and the corresponding dimeric derivatives.

2.7 Deprotection reactions

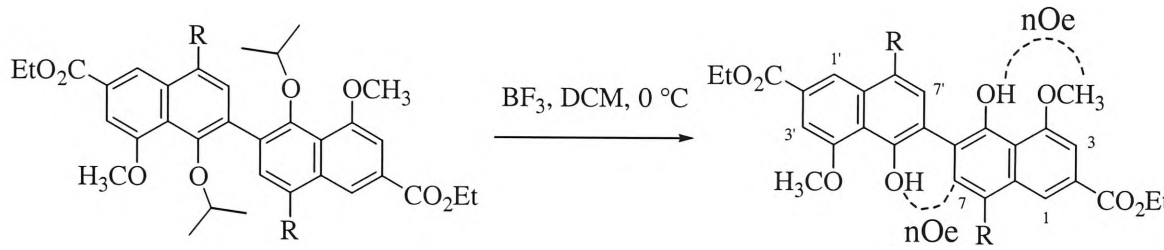
The final reaction in the synthesis of the targeted compounds of Figure 2.2 was the deprotection of the isopropoxy group, to yield the hydroxy groups *ortho* to the biaryl axis. These hydroxy groups were a key feature of the inhibitor design of Section 2.1.1. In a typical reaction, boron trifluoride diethyl etherate was added to a stirred solution of the isopropoxy protected compound in DCM, at 0 °C under nitrogen (Scheme 2.30). The reaction solution was normally stirred for two hours before the addition of methanol to quench the reaction. Boron trifluoride was chosen preferentially over boron trichloride to specifically cleave the isopropyl group, and not affect the methoxy group, or the ethyl ester. However this precaution may have affected the yields of Table 2.12 that show the moderate yields for this reaction. A boron trichloride deprotection by Bringmann *et al.* of a compound similar to the brominated naphthyl (**65**) used for the cross-coupling reactions, but with a methyl substituent instead of the ethyl ester, was reacted for only 15 minutes and yielded 91 % of the deprotected product.[62]

Scheme 2.30



R = phenyl **(68)**
naphthyl **(72)**
pyrimidine **(83)**
methyl **(79)**

R = phenyl **(99)**
naphthyl **(100)**
pyrimidine **(101)**
methyl **(102)**



R = phenyl **(91)**
naphthyl **(96)**
methyl **(97)**
pyrimidine **(98)**

R = phenyl **(103)**
naphthyl **(104)**
methyl **(105)**
pyrimidine **(106)**

R	monomer	dimer
phenyl	(**99**) 64 %	(**103**) 68 %
naphthyl	(**100**) 89 %	(**104**) 68 %
pyrimidine	(**101**) 56 %	(**105**) 65 %
methyl	(**102**) 55%	(**106**) 71 %

Table 2.12 Yields of the deprotected monomer and dimer target compounds, Scheme 2.30.

The ^{1}H NMRs of the deprotected compounds typically saw the loss of the doublet of the isopropyl methyl groups that was usually about δ 1.44 in the monomer compounds, and δ 1.10 in the dimer compounds. The quintet of the isopropyl methine hydrogens at about δ 4.59 in the monomer compounds and δ 4.17 in the dimer compounds was also absent in the deprotected compounds. In the deprotected phenyl substituted monomer (**99**), NOESY experiments detected an nOe between the naphthol at δ 9.49 and the methoxy at δ 4.45 (Scheme 2.30). In the deprotected dimerised phenyl product (**103**), a similar nOe was seen between the naphthol in one naphthyl system at δ 9.99 and the

methoxy in the same aromatic system at δ 4.14. Another nOe was seen between the naphthol and the 7, or 7' hydrogen of the opposite naphthyl system at δ 7.59.

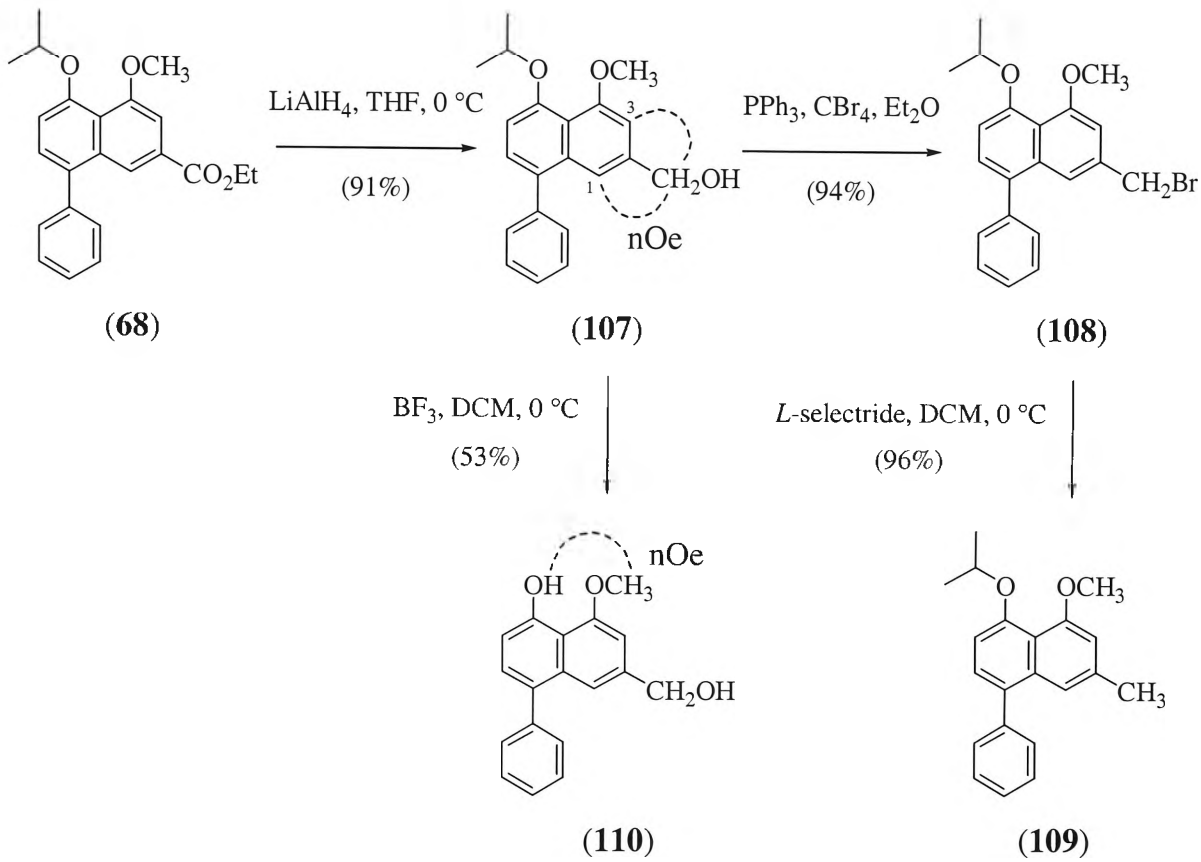
The loss of the isopropyl group, typically had a small downfield shifting affect on the methoxy groups of all of the deprotected compounds of Scheme 2.30, of the order of δ 0.11. The naphthyl hydrogens of the monomeric series of products generally showed only very small changes in their chemical shifts, and varied in direction between the different R-substituents. The only notable large shift was in the deprotected methyl monomeric product (**102**), which saw a downfield shift in the 3-H from δ 7.31 in (**79**) to δ 8.33 in (**102**). The assignment of 1-H and 3-H was confirmed by an HSQC experiment, and a NOESY experiment indicating an nOe between the 3-H and the methoxy substituent. The naphthyl hydrogens of the deprotected dimer series of compounds, (**103**) – (**106**), showed consistent downfield shifts for the 1 and 1' naphthyl hydrogens, and consistent upfield shifts for the 3, 3', 7', and 7' hydrogens. These shifts were generally small, except for those of the 7 and 7' hydrogens that shifted upfield, on average by δ 0.38.

2.8 Transformations of the ethyl ester

With the greater part of the synthesis of the monomeric and dimeric series of targeted compounds completed (Figure 2.2), preliminary studies were conducted to transform the ethyl ester sidechain of these compounds (Scheme 2.31). The lithium aluminum hydride reduction in THF transformed the ethyl ester to a primary alcohol functionality, yielding (**107**) in 91 % yield. This was comparable to the 93 % yield reported by Bringmann *et al.* for a similar reaction, in which the phenyl substituent of (**107**) is replaced by a bromine.[53,62] The mass spectra of the methyl alcohol derivative (**107**) revealed an *M*+1 at 323, and an *m/z* at 281, which is likely to be the loss of the isopropyl group. The ^{1}H NMR also shows the loss of the indicative triplet and quartet at δ 1.34 and δ 4.34 of the ethyl ester (**68**), and the appearance of a singlet of 2H integration at δ 4.67, corresponding to the methylene of the primary alcohol (**107**). Small upfield shifts were seen in the naphthyl hydrogens due to the shielding effects of the primary alcohol, or rather due to the loss of the deshielding effects of the ethyl ester. The most notable of these were in the adjacent 1-H and 3-H that shifted from δ 8.22 to δ 7.34 and δ 7.48 to δ 6.89, respectively, compared with (**68**). A NOESY experiment also detected nOes

between the methylene hydrogens of the primary alcohol at δ 4.67, and the 1-H and 3-H hydrogens, as seen by the dotted line in (**107**) of Scheme 2.31.

Scheme 2.31



To achieve the methyl group at the 2 position that is analogous to the substituent in the michellamine structure (**26**), the corresponding methyl bromide (**108**) was first needed. Attempts to follow the methodology of Bringmann *et al.* and use 1,2-dibromotetrachloroethane and triphenylphosphine in dichloromethane gave no product (**108**) and only starting material was recovered.[53,62] An alternate synthesis with carbon tetrabromide and triphenylphosphine in ether[117] afforded the bromide (**108**) in 94 % yield, comparable to the 93 % yield achieved by Bringmann *et al.* using 1,2-dibromotetrachloroethane. The mass spectra of (**108**) revealed *m/z* peaks at 385 and 387 corresponding to *M*+1 peaks for the bromine isotopes of 79 and 81 respectively. Mass to charge peaks of 343 and 345 were identified as the loss of the isopropyl group and a *m/z* at 307 corresponded to the further loss of the bromine. The methylene hydrogens at the 2 position were shielded by the bromine and shifted upfield to δ 4.50 from δ 4.67 in the ^{1}H NMR of (**107**). ^{13}C NMR and DEPT experiments also identified δ 34.64 as a methylene carbon, which HSQC experiment correlated to the δ 4.50 signal.

The methodology of Bringmann *et al.* was used in the final reaction to realise the 2-methyl group.[53,62] Two successive aliquots of a solution of *L*-selectride in THF were added to a solution of the bromide (**108**) in DCM, followed by the addition of ethanol to quench the reaction. The resulting product (**109**) was isolated in 96 % yield. The mass spectra of (**109**) identified an *M*+1 at 307, and a *m/z* at 265, indicative of the loss of the isopropyl group. ^{1}H NMR showed a change in the integration and shift of the two methylene hydrogens at δ 4.50 in (**108**) to a signal of three hydrogen integration at δ 2.36 that was identified as the aryl methyl. All of the naphthyl hydrogens showed upfield shifts, most notably in the adjacent 1-H and 3-H that shifted from δ 7.39 to δ 7.19 and δ 6.86 to δ 6.70, respectively compared to the methylene bromide derivative (**108**).

The primary alcohol derivative (**107**) was treated with boron trifluoride to remove the protection of the isopropyl group and yield the deprotected methyl alcohol derivative in 53 % yield (**110**). The mass spectra of (**110**) identified an *M*+1 at 281. The ^{1}H NMR lacked the characteristic doublet of the isopropyl methyl hydrogens at δ 1.44 in the protected species (**107**), as well as the quintet of the isopropyl methine hydrogen at δ 4.58. The naphthyl hydrogens showed varying shift changes; 1-H and 7-H were shifted slightly downfield, and 3-H and 6-H were shifted slightly upfield. Like the deprotected compounds of Scheme 2.30, the methoxy hydrogens of (**110**) saw a downfield shift from δ 3.98 to δ 4.08 compared to (**107**). A NOESY experiment also detected an nOe between the naphthol at δ 9.48 and the methoxy at δ 4.08.

The boron trifluoride deprotection of (**109**), and the further investigation of the transformations in the ethyl esters of all the R-substituted products in both the monomer and dimer series of targeted compounds (Figure 2.2) were not completed by the conclusion of this research. Their completed synthesis will play an important role in increasing the structural diversity of the targeted compounds, and providing more variations and derivatives for biological testing. The results from these compounds will provide further information to help form the pharmacophore for antiviral and antimalarial activity, and additional aid in future inhibitor design.

2.9 Antiviral and antimalarial testing

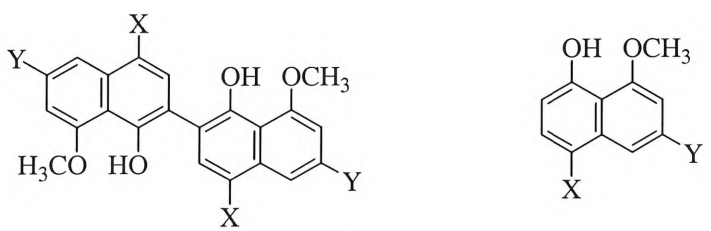
The deprotected compounds of Scheme 2.30 and Table 2.12 were screened for antiviral and antimalarial activity. The results of these tests were not yet available at the time of printing.

Antiviral testing is currently being carried out through AMRAD Operations[118], using a standard reverse transcriptase assay (HTLV-III_B infected H_9 cell line) involving incorporation of labelled dTTP into a primer and then measuring incorporated radioactivity. Nevirapine is used as a standard in this assay.

Antimalarial testing is currently being carried out through Professor Yodhathai Thebtaranonth, using a World Health Organisation certified protocol, detailed in.[119]

2.10 Future synthetic targets

The completion of the synthesis of the dimeric and monomeric series of designed inhibitors, illustrated in Figure 2.12, remains the leading objective of this research. The establishment of the methodology for the Suzuki cross-coupling and homo-coupling reactions in this research should facilitate the completion of this initial phase of the inhibitor synthesis presently. Analogues with transformations of the ethyl ester sidechain, illustrated as the Y substituent in Figure 2.7, were also intended for synthesis and these should remain important synthetic targets for inhibitor design. The preliminary trial reactions illustrated in Scheme 2.31 (Section 2.8) have established that the methyl alcohol and methyl substituents can readily be synthesised from the ethyl ester.



X = phenyl	Y = COOEt
naphthyl	COOH
methyl	CH_2OH
pyrimidine	CH_3
pyridine	

Figure 2.7 Dimeric and monomeric series of inhibitors for forthcoming synthesis.

The carboxylic acid substituent could be transformed using an acid or base-catalysed ester hydrolysis reaction. These transformations, in each of the X-substituted inhibitors, for both the dimeric and monomeric series, will introduce significant and necessary structure-activity relationship (SAR) data once synthesised and tested for antiviral and antimalarial activity. In particular, inhibitors in which the Y-substituent is a methyl group should provide valuable SAR data, as the michellamine alkaloids, from which most of the inhibitor design was derived, have a methyl substituent in this position (Figure 2.1).

The directions for future inhibitor design will depend very much on the results of the antiviral and antimalarial testing. The scope for possible analogues, with different X- and Y-substituents is endless, and the selection of analogues for synthesis should be guided by rational drug design principles, based on the SAR data obtained from current biological testing. Another consideration that should influence future inhibitor design is the incorporation of features from other naturally derived antiviral agents. A hydroxy group, a prominent substituent of gossypol, the tetrahydroisoquinoline unit of the michellamines (**26**) (Figure 2.1), and many of the naturally derived antiviral agents of Figure 1.7 could also replace this methoxy group to add diversity to the designed inhibitors for SAR data.

Chapter 3

Conformational Studies of Reverse Transcriptase

3.1 Reverse transcriptase

The discovery of reverse transcriptase (RT) in 1970[120,121] set a new paradigm for the flow of genetic information from RNA to DNA, and established a genetic and biochemical basis for the identification of retroviruses. Reverse transcription flagrantly violates the central dogma of molecular biology: that the flow of genetic information was unidirectional, from DNA to RNA to protein.[122] Reverse transcription by HIV-1 RT converts single-stranded genomic RNA (ssRNA) into double stranded DNA (dsDNA) for integration into the host chromosome. The pivotal and unique role RT plays in the life cycle of the virus has made it a target for anti-HIV drug therapy.[26,27] Reverse transcriptase has also had a tremendous impact as a biotechnological tool for the synthesis and cloning of DNA from messenger RNA.[123]

3.1.1 Structure of HIV-1 reverse transcriptase

RT is a multifunctional enzyme, whose structure, function and mechanism has been extensively studied and reviewed.[22,122,124-128] This enzyme possesses both RNA-dependent and DNA-dependent DNA polymerase activities, as well as a ribonuclease hybrid (RNase H) activity that specifically degrades the RNA strand of RNA/DNA hybrids. RT also has an RNA/RNA duplex-dependent ribonuclease activity termed RNase H*.[129]

In mature HIV-1 viral particles, RT exists as a highly asymmetric heterodimer consisting of 66 and 51 kDa subunits. The amino terminal of the two subunits is

identical. The p51 subunit is derived by proteolytic processing of the p66 subunit between Phe440 and Tyr441 by HIV-1 protease and lacks the RNase H domain.

The catalytically active p66 subunit is folded to form an 'open right-hand' configuration, creating a large DNA-binding cleft in the polymerase site that exposes the catalytic residues.[130] The anatomical resemblance led to subdomains termed 'fingers', 'thumb', and 'palm'. A 'connection' subdomain lies between the palm and the RNase H subdomain located at the carboxy-terminal end of p66 (Figure 3.1, Appendix 7.1 Subdomains of reverse transcriptase). The p66 subunit also contains the binding pocket for non-nucleoside reverse transcriptase inhibitors.

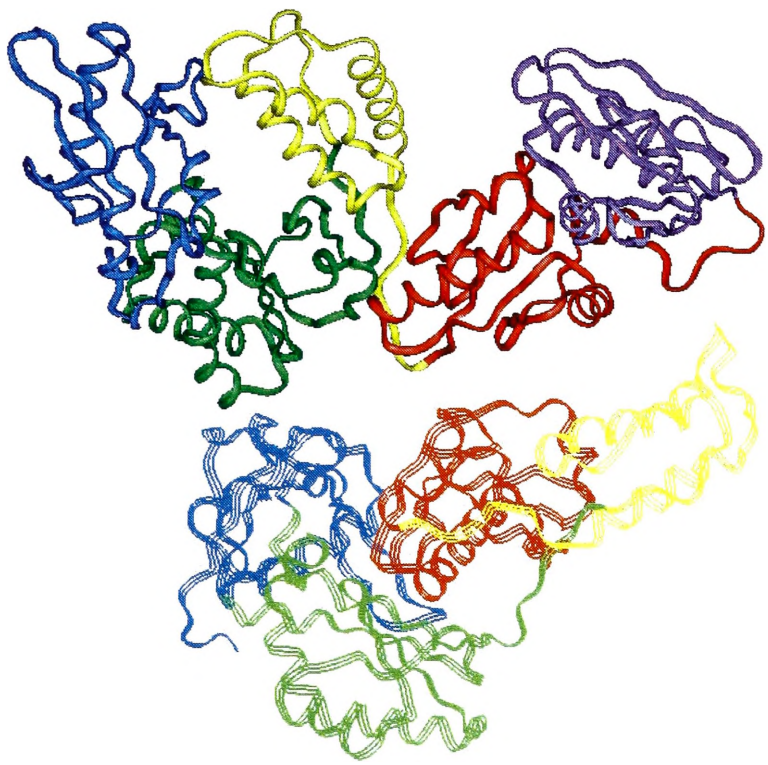


Figure 3.1 Structure of HIV-1 reverse transcriptase heterodimer. The subdomains of the p66 (ribbon) and p51 (lines) with their disparate folding and packing; fingers (blue), palm (green), thumb (yellow), connection (red), and RNase H (purple).

Although the amino-terminal sequence of both subunits is identical, p51 does not directly contribute to the various catalytic activities. The subdomains of p51 are more tightly packed and are folded very differently to those of the p66. The thumb is rotated away from the fingers, which are closed over the palm in a manner that buries the catalytic residues, and the connection is folded onto the palm between the fingers and the thumb subdomains. The p51 subunit is catalytically inactive, but interacts with the

RNase H domain in the heterodimer, and may have a role in altering the processivity of the p66 subunit, facilitating strand displacement, DNA synthesis, tRNA binding, as well as providing scaffolding support to the p66 subunit.[131]

The secondary structural classification of RT has been determined by two different groups: the Center for Advanced Biotechnology and Medicine (CABM) at Rutgers University and the Oxford Centre for Molecular Sciences (OCMS) at Oxford University.[132,133] The nomenclature for secondary structures used throughout will be that of the CABM research group (see Appendix 7.2 Secondary structure of reverse transcriptase).

3.1.2 Reverse transcription

The reverse transcription of viral RNA to proviral DNA (Figure 3.2) is made possible by an elegant and complex mechanism that has been extensively studied and examined to better understand its workings, and ultimately prevent its completion.[5,122,127,128] Reverse transcription is a discontinuous process occurring in the viral nucleocapsid in the presence of an excess of enzyme with respect to template. Between 30 and 50 RTs are present in each particle, for the two RNA genomes. This excess of enzyme is attributed to the necessity to overcome the low processivity of the enzyme; excess enzyme ensures recycling of RT on nucleic templates, enhancing the probability of obtaining completed reverse transcripts.[22]

Step 1 - Initiation of reverse transcription - Binding of the tRNA primer to the primer binding site

All retroviruses utilise host-encoded transfer RNA (tRNA) species as primers for reverse transcription. HIV-1 uses $tRNA^{Lys3}$, which anneals to the viral (+) strand RNA at a region, approximately 200 nucleotides downstream from the 5'-end of the RNA, known as the primer binding site (PBS). The host tRNA is partially unfolded from its native structure so that the 18 nucleotides at the 3'-terminal end of human $tRNA^{Lys3}$ that are complementary to the PBS can base pair to form a binary tRNA/RNA complex. This complex is recognised by RT, which then initiates reverse transcription by extending the 3'-end of the annealed primer.

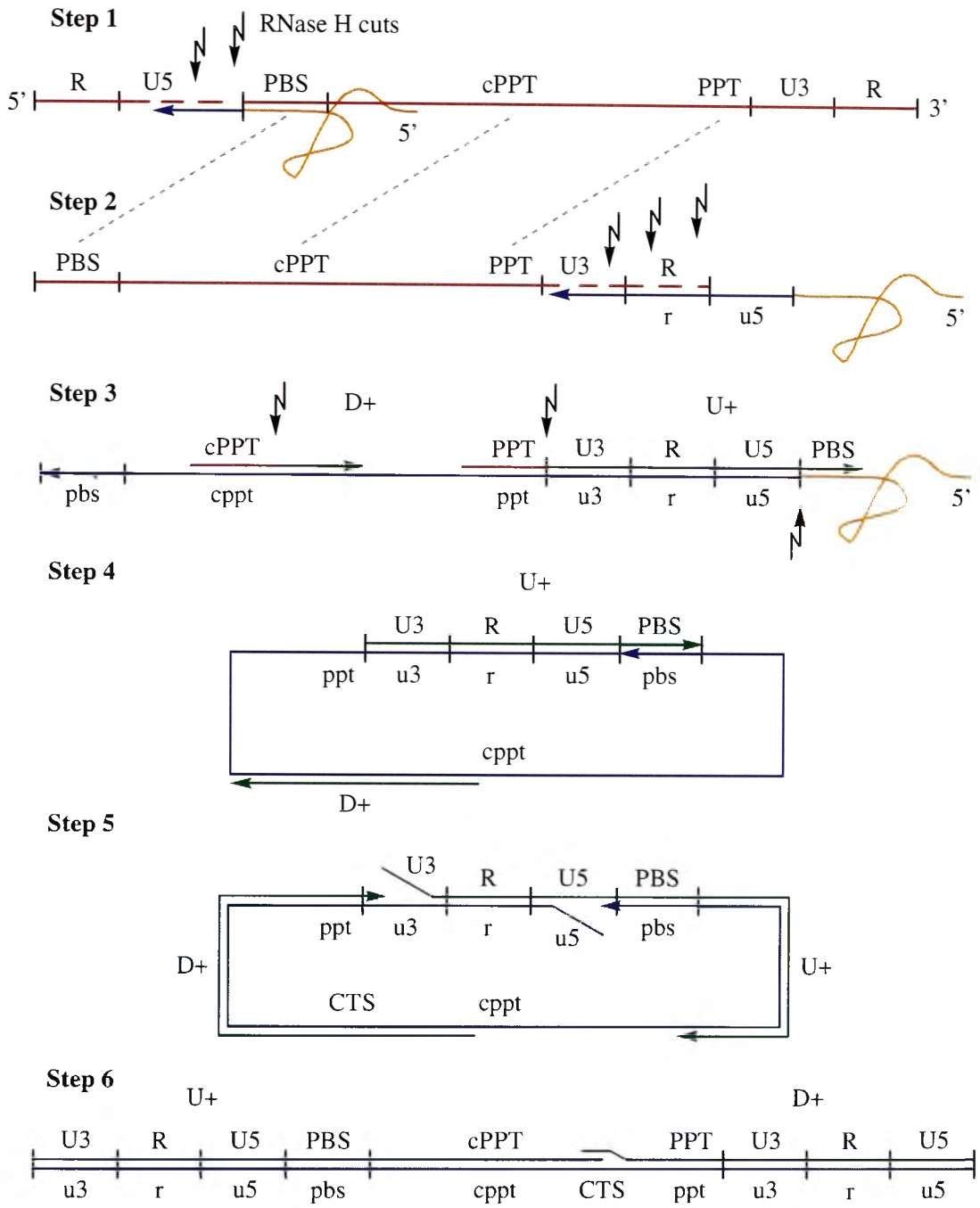


Figure 3.2 The process of reverse transcription. Red lines represent RNA strands, blue lines represent (-) strand DNA, and green lines represent (+) strand DNA. The curved orange line represents the tRNA primer. Arrows represent the elongating DNA strand. Uppercase letters indicate (+) strand regions, and lowercase letters represent (-) strand regions. Abbreviations are explained in the text. Adapted from Götte *et al.*[127]

Synthesis of (-) strand DNA is initiated at the 3'-hydroxyl end of the $tRNA^{Lys3}$ with the genomic RNA serving as a template and proceeds toward the 5'-end of the RNA template. Concurrent with DNA synthesis, the RNA strand of the newly formed RNA/DNA hybrid is digested by the RNase H domain of RT. RNase H is able to discriminate between DNA/RNA and RNA/RNA substrates, and initially degrades the

RNA strand upstream, towards the 5' end, of the PBS. The PBS sequence which formed an RNA/RNA homoduplex with $tRNA^{Lys3}$ to initiate reverse transcription remains resistant to RNase H degradation at this stage.

Step 2 - First strand transfer

When RT reaches the 5'-end of the viral RNA template, the DNA transcribed thus far is referred to as minus-strand strong-stop DNA, (-)ssDNA. In order to complete the synthesis of the (-) strand DNA, the (-)ssDNA must be transferred to the 3'-end of the genomic RNA. This strand transfer is facilitated by the terminal repeat (r) region of the nascent (-)ssDNA, that is complementary to the R region of the 3'-end of the RNA template. RNase H degradation of the 5'-end of the RNA releases the (-)ssDNA, allowing its translocation to the R region of the 3'-end of the genomic RNA. Since viral particles contain two homologous copies of genomic RNA, the first strand transfer may occur either as an intra- or intermolecular event.

Step 3 - Synthesis of (+)-strand strong-stop DNA - Initiation by purine-rich primers

As synthesis of (-) strand DNA continues, specific RNase H degradation of the RNA template creates purine-rich fragments that function as RNA primers for (+) strand DNA synthesis. These regions of the RNA template that are highly resistant to RNase H degradation are located close to the genomic 3'-end and in the centre of the viral RNA, and are referred to as the polypurine tract (3'PPT) and the central PPT (cPPT) respectively. Before (-) strand DNA synthesis is completed, the synthesis of two (+) strand DNA segments will have been initiated by these primers, with the nascent (-) strand DNA as template. The two distinct (+) strand DNA segments initiated by the cPPT and the 3'PPT are called downstream (D+) and upstream (U+) segments respectively. As the U+ DNA synthesis encounters the tRNA sequence of the $tRNA^{Lys3}$ primer still attached to the 5'-end of the (-) strand DNA, the DNA transcribed is referred to as plus-strand strong-stop DNA, (+)ssDNA. This reverse transcription of the 3'-end of the tRNA regenerates the PBS sequence at the 3'-terminus of the (+)ssDNA.

(+)ssDNA synthesis continues until the first modified tRNA base, a methylated adenosine, is encountered 19 residues upstream from the 3'-end of the $tRNA^{Lys3}$. Specific RNase H cuts then allow the removal of the $tRNA^{Lys3}$ and PPT primers from the (+) and (-) strand DNA respectively.

Step 4 - Second strand transfer

The second strand transfer is necessary for (+)ssDNA and U+ DNA syntheses to continue, and is facilitated by the 18 nucleotide PBS. At this stage of reverse transcription, there are two copies of the PBS; the first is the (-) strand DNA copy that is formed when (-) strand DNA synthesis advances into the PBS region of the genomic RNA, and the second is the (+) strand DNA copy that is regenerated by using tRNA as a template during the synthesis of (+)ssDNA (Step 3). The PBS facilitates the second strand transfer in what is presumed to be an intramolecular event involving the formation of a circular intermediate. The signal for pausing of (-) strand DNA synthesis is the first RNase H cut at the tRNA-DNA junction that was formed during initiation of (-) strand DNA synthesis.

Step 5 – Strand displacement and continuation of DNA synthesis

To ensure that the entire viral genome is transcribed, RT has a strand displacement activity that disrupts DNA/DNA duplexes. This allows continuation of the syntheses of (-) strand DNA and D+ DNA at the U3-R-U5 duplex. The U+ DNA segment serves as a template to complete the (-) strand DNA, while the u3 segment serves as a template for the completion of the synthesis of the D+ segment.

Step 6 - Completion of transcription and termination

Synthesis of the U+ segment continues until RT reaches the central terminal sequence (CTS), located at the centre of the genome. RT displaces part of the 5'-end of the D+ DNA segment, allowing U+ DNA synthesis to terminate 100 nucleotides downstream of the 5'-end of the D+ segment. The resulting double stranded DNA carries a (+) strand overlap at its centre. The legacy of the two strand transfer reactions is evident in the provirus' acquisition of duplicated U3-R-U5 sequences at both ends, known as long terminal repeats (LTR). The LTRs are critical for the integration of the viral DNA into the host genome. The final dsDNA product of reverse transcription, called proviral DNA, is longer at both ends than the original template viral RNA, and ready to be integrated into the host cell genome.

The unique replicative mechanism of reverse transcription has made reverse transcriptases one of the most targeted viral processes in the life cycle of HIV for drug inhibition. Its characteristic structure and multiple functions provide numerous sites and steps in the catalytic cycle for inhibition. While nucleoside analogues were the first

drugs developed to arrest the polymerisation reaction of DNA synthesis, it was the advent and evolution of non-nucleoside inhibitors that have lead efforts to find specific and efficacious drugs against RT.[24] Virtually all of these inhibitors belong to the class of non-nucleoside reverse transcriptase inhibitors (NNRTIs) which bind at a site distinct from the polymerase active site of the enzyme, and exert their inhibitory influence on the polymerisation reaction by allosteric means.[26,27]

3.1.3 The non-nucleoside inhibitor binding pocket (NNIBP) and inhibition of RT

The inherent flexibility of HIV-1 RT is of manifest significance in the execution of its multiple biological functions. This flexibility accommodates the binding of the template-primer complex, inducing numerous conformational changes throughout the heterodimer. These changes are believed to provide a means of allosteric communication between spatially distant regions such as the polymerase and RNase H active sites, during the potentially coupled processes of dNTP incorporation and RNA template degradation.[134] This flexibility is no more evidently demonstrated than in the gross displacement of the thumb subdomain from a position nearly touching the fingers subdomain to an upright configuration, upon binding of the nucleic acid substrate (Figure 3.3).[132,135] Other notable conformational changes include displacements in sections of the fingers domain which 'flex' towards the palm, playing a possible role in allowing the release of pyrophosphate after a polymerisation reaction, translocation, and binding of the next dNTP during DNA elongation.[136]

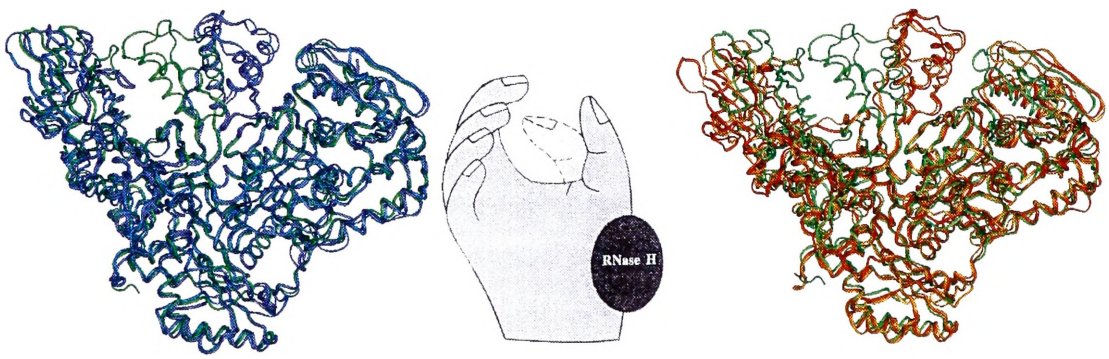


Figure 3.3 Ribbon illustrations of the reverse transcriptase enzyme showing movement of the thumb subdomain, approximately 30 Å, upon the binding of nucleic substrates or non-nucleoside inhibitors. The thumb in the unliganded enzyme (1dlo-green) comes almost into contact with the fingers in a 'closed' conformation. (Left) The DNA bound structures (1rtd-dark blue and 2hmi-light blue) show an extension of the thumb in an 'open' conformation. (Right) Inhibitor-bound structures (1hni-orange and 1vru-red) have the thumb subdomain further extended in the 'open' conformation. Diagram made using superimposition techniques detailed in Section 3.2.

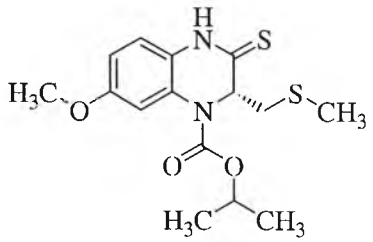
Ironically, the very flexibility essential to the successful performance of the enzyme is also fundamental in its inhibition by non-nucleoside inhibitors. NNRTIs are known to bind in a hydrophobic pocket in the palm subdomain, some 10 Å from the polymerase active site. In the absence of an inhibitor, the pockets exists as a mere surface depression, with the sidechains of two important tyrosine residues pointing into the hydrophobic core.[30] The bulk of the inhibitor is accommodated for by a differential twisting of the β12-β13-β14 sheet that results in an expansion of the pocket. This results in a shift of the β6-β9-β10 sheet that contains the catalytic residues of the polymerase active site to a conformation similar to that observed in the inactive p51 subunit.[29] While this conformational change may be sufficient to explain how the chemical step of phosphodiester bond formation is blocked, it does not account for the entirety of NNRTI inhibition. NNRTIs binding exerts varying effects on several RT features including, but not limited to, enzyme processivity, binding of different template-primers and sequences, recognition of different nucleic acid substrates, and cleavage specificity of RNase H activity.[79]

Several hypotheses have been proposed to give a rational interpretation of the mechanism, or mechanisms, of non-nucleoside inhibition.[79] Of the three dominant inhibition mechanisms, the 'active site distortion model' proposes that conformational changes in the NNIBP distort the precise geometry, mobility, electrochemical environment, and solvent accessibility of the polymerase active site and its coordination geometry with divalent metal ion complexes.[79,134] This distortion cripples the enzyme's ability to interact with constantly changing base pairs and template primer substrates, distinguish different dNTP substrates, or permit translocation of the template-primer following nucleotide incorporation. The 'primer grip model' postulates that the creation of the NNIBP significantly displaces the primer grip, which is composed of the β12-β13 hairpin. The displacement of the primer grip may effect the position and mobility of the primer terminus, or alter the stereochemical relationship between the primer terminus, a bound dNTP, and associated divalent metal cations at the polymerase active site.[79] The final 'molecular arthritis model' speculates that NNRTI binding leads to a restriction in the mobility of the p66 thumb subdomain which, by virtue of its extensive interactions with the template-primer, may interfere with the processivity of polymerisation.[130,137] These models are not mutually exclusive; no model on its own is sufficient to explain

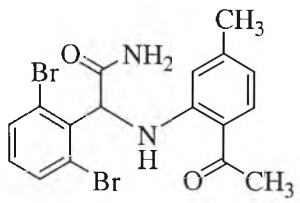
all the kinetic and biochemical evidence of non-nucleoside inhibition. Rather it is believed that several mechanisms are involved in the allosteric inhibition process.[79]

The study of these mechanisms of inhibition has, for the most part, been qualitative and based on general observations. Those few studies that have attempted to quantify the measure of inhibition have not always been impartial, considering only structures solved in their own laboratories, or wholly rational in their approach, considering only a few experimental structures at a time.[133,138,139] Our research endeavored to resolve the lack of objective, quantitative analyses of the conformational changes of several structurally significant active sites and motifs of RT, and their role in the mechanisms of inhibition upon binding of NNRTIs. The conformational changes in the residues of the polymerase and RNase H active sites, along with the primer and template grips and the NNIBP were scrutinized to ascertain their role in inhibition.

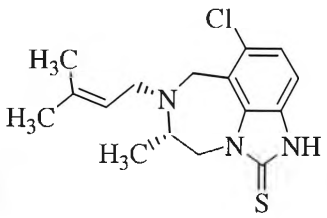
The wealth of interest in RT has generated a large number of x-ray crystallographic structures that allowed both inhibitor and substrate induced conformational changes in RT to be investigated. RT structures were available in the unliganded form, complexed with a dsDNA template-primer and a Fab monoclonal antibody, a dsDNA template-primer and a dNTP, and a wealth of structurally diverse inhibitors (Appendix 7.3 X-ray crystallographic structures of reverse transcriptase). These inhibitors (Figure 3.4) included the clinically used nevirapine (**8**), and tivirapine (**11**) and MKC-442 (**12**) in clinical trials. This study considers a total of 23 crystal structures from both of the major research groups operating in the field.



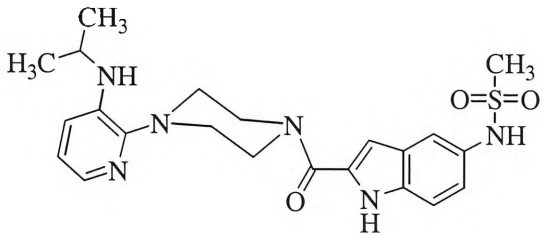
HBY097 (**111**)



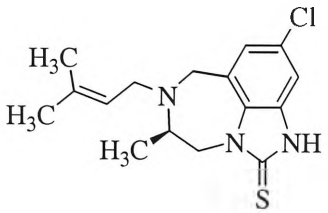
Br-α-APA (**112**)



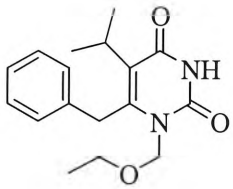
TIBO R86183 (**11**)



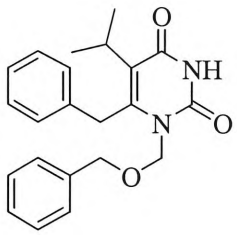
BHAP U90152 (**113**)



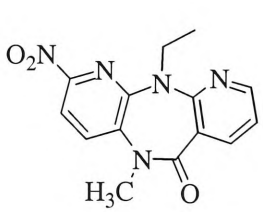
TIBO R82913 (**114**)



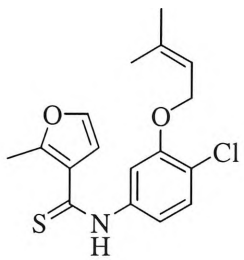
MKC-442 (**12**)



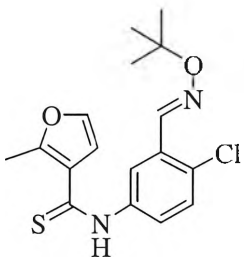
TNK-651 (**115**)



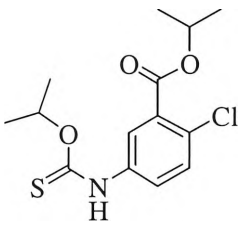
1051U91 (**116**)



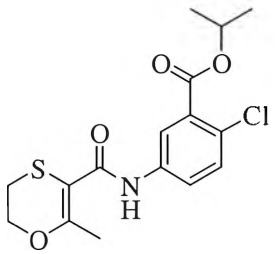
UC781 (**117**)



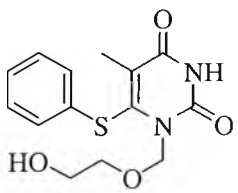
UC10 (**118**)



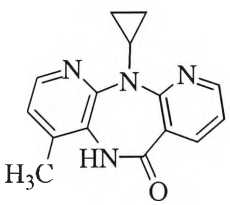
UC38 (**119**)



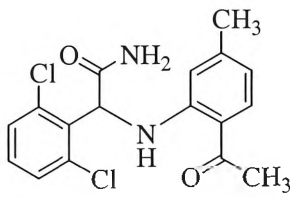
UC84 (**120**)



HEPT (**121**)



nevirapine (**8**)



Cl-α-APA (**122**)

Figure 3.4 Structures of non-nucleoside inhibitors from x-ray crystallographic RT-inhibitor complexes (see Appendix 7.3 X-ray crystallographic structures of reverse transcriptase).

3.2 Superimposition

The dynamic nature and design of proteins have long been a source of unanswered questions and challenges for all fields of science, in particular the mechanics of protein function and how these processes can be disrupted. The structure comparison of different conformations of a protein can provide insight into these fields of research, and help in the optimisation of current drugs as well as the *de novo* design of structurally unique inhibitors.

The traditional method for the analysis of two related structures is to perform an optimal rigid body superimposition, giving equal weighting to all regions of the structure. The similarity of the two structures is customarily expressed as a root-mean-square deviation (RMSD) of the atomic coordinates used in the superimposition.[140] The superimposition of the entire backbone, or all C_α atoms, of a protein runs the inherent and probable risk that small differences between the two proteins will be averaged out and missed.[141] This effect would be particularly pronounced in flexible protein systems such as RT. The use of subsets other than the backbone, or all the C_α atoms, has been contentious, with a number of schools of thought regarding the best regions of a biomolecule to choose for superimposition. One method is to select atoms directly surrounding the region to be examined, either by choosing a sequence of residues[29,139] or by creating a subset of residues within a given radius of the region to be examined.[31,142] Other methods involve the use of polypeptide sections at either end of the region to be examined,[143] or a 'core' region of the protein.[133,135,138] The rationale behind the selection of atoms to be used for superimposition has not always been made clear, and is not consistent amongst different authors; in one example, two different superimposition techniques were used within the same research paper to accentuate different points.[144] Results from these studies are vulnerable to error and misinterpretation as a result of user-bias introduced in selecting those atoms to form the superimposition subset. Different results can be obtained depending on the atoms selected for the superimposition (see Section 3.2.3).

3.2.1 Difference distance matrices

In order to investigate conformational changes in flexible protein systems, we needed to define a superimposition strategy that provided a definitive, unambiguous and objective

procedure for the superimposition of large molecules. To achieve this, we investigated the use of difference distance matrices (DDMs). DDMs allow the rapid identification of regions of conformational similarity and dissimilarity, and from these, suitable rigid regions can be selected for superimposition studies.[145,146] They are generated by the comparison of the distance matrices (DMs) of two different molecules.

DMs are constructed by calculating the distances (r_{ij}) between the i^{th} and j^{th} atoms, for every combination of two atoms in a single molecule or protein. These distances are represented in an n x n square matrix (n equal to the number of atoms in the molecule), with the atoms i and j along the axes. In this internally referenced matrix, pairs of atoms that are close to each other in three-dimensional space have small values of r_{ij}, while those that are distant have larger r_{ij} values. Due to the sizeable number of atoms in most proteins, only the C_α atoms of each residue of the protein are considered in the DM calculation.[147]

The DDM, comparing two protein conformations A and B, is constructed by first calculating separate DMs containing the distances r_{ij}^A and r_{ij}^B from conformations A and B respectively. One DM is subtracted from the other, by calculating the difference (Δr_{ij}) between elements r_{ij}^A and r_{ij}^B that have corresponding values of i and j.[145]

$$\Delta r_{ij} = |r_{ij}^A - r_{ij}^B| \text{ for all combinations of } i \text{ and } j$$

The DDM is the representation of this data in an n x n square matrix with i and j along the axes.

If two protein conformations A and B are identical, the corresponding distances r_{ij}^A and r_{ij}^B will be identical and the difference Δr_{ij} equal to zero. However, if a residue x is in a different position in one structure, it will have different distances to every other residue j in the protein, with all Δr_{ij} values being nonzero where $i = x$. Consequently, regions of conformational dissimilarity between the two structures will appear as clusters of nonzero difference values in the DDM, with the value of Δr_{ij} indicating the magnitude of conformational change.[147]

DDMs have previously been used for the direct analyses of movement or in the creation of subsets for subsequent studies in biological systems such as Human Cardiac Troponin C,[148] human hemoglobin,[149] and calbindin D_{9k}.[150] Jäger *et al.* have previously used DDMs to a limited degree to assess which portions of the RT enzyme behave as

rigid bodies in different crystal lattice environments.[143] They concluded that the RT enzyme has a specific flexibility that allows rotation of the polymerase active site relative to the rest of the molecule, speculating that this swivel motion may be involved in the translocation of the growing duplex after incorporation of a new nucleotide.

In our studies, DDMs were used to identify polypeptide regions of RT which were conformationally similar in unliganded, inhibitor-bound, and substrate-bound x-ray crystallographic structure, in an unbiased and impartial manner. These regions were superimposed to allow the movement of the localised variable regions, which were not used in the superimposition calculation, to be readily identified with confidence, and accurately quantified. Thus, comparison studies of RT could be undertaken with the unique demands of proteins, when studying their movement and flexibility, taken into account.

3.2.2 Derivation of the superimposition subset

The ProFlex program was developed in our research group to calculate DMs and DDMs of different proteins and protein conformations.[151] This was used to identify a subset of C_α atoms within RT that did not move upon binding of an inhibitor, a substrate, or compared to another unliganded x-ray crystallographic structure.[147,151] Initially the superimposition subset was derived from the DDMs of 19 liganded and unliganded crystal structures compared against 1dlo (all crystal structures from Appendix 7.3, excluding 1hmv and 1rt4-1rt7). The 1dlo crystal structure was chosen as the unliganded RT crystal structure over 1hmv and 1rtj due to its higher resolution and data completeness. The structure of 1rtj was ruled out because of known irregularities in the crystal collection method;[79] its crystal structure was prepared by soaking out the non-nucleoside inhibitor HEPT from a HEPT/RT complex.[29] The resulting crystal structure was not significantly different from the inhibitor-bound structure, as if the enzyme had not been given sufficient allowance to return to a native unliganded conformation.

As the crystal structures had varying, missing or undefined polypeptide regions, the 20 crystal structures, including 1dlo, were aligned and only those amino acid residues present in all 20 structures were translated to DMs. This ensured that each crystal structure contributed equally to the calculations. Only the C_α coordinates were used in these calculations, as the inclusion of all the atom coordinates would have been logistically infeasible. From these DMs, 19 DDMs were calculated comparing all

crystal structures to that of 1dlo. In each DDM, any residue with a calculated difference distance between its two conformations of less than 2.0 Å was defined as being conformationally invariable. 2.0 Å was chosen as smaller or larger values yielded too few and too many residues, respectively, for optimal superimposition and RMSD values.[151] The final superimposition subset, listed in Table 3.1, was restricted to only those conformationally invariable residues that were common to all 19 crystal structures, ensuring that anomalies in any one structure were not included.

subunit	residues
p66	4-6, 95-107, 162-163, 180-181, 188-200, 202-205, 226, 234-235, 237-239, 317, 319, 323, 339-345, 349-353, 365-366, 368-402, 405-419, 428-436, 439, 493, 530
p51	6-7, 18-45, 54-64, 71-84, 97-111, 113-117, 121, 123-138, 140-174, 176-184, 186-192, 197-198, 201-202, 208, 252, 254-264, 267, 274, 277, 280-282, 284, 296, 298-300, 303-307, 320, 322, 329, 331, 333-355, 364-393, 397-417
total : 376 residues - 1528 atoms (backbone atoms of 376 residues)	

Table 3.1 List of the residues of the DDM derived superimposition subset.

The DDMs of 1hmv and 1rt4-1rt7 compared to 1dlo were calculated in the same manner at a later stage, after the public release of the coordinates for 1rt4-1rt7. The superimposition subsets derived from these DDMs were found to be compatible with the original subset, further validating the invariability of the residues of the subset, and were used in the superimposition studies. To maintain the integrity of the superimposition subset and procedure, residues that were of interest to be studied for their role in the inhibition of RT were not removed from the superimposition subset. The final subset (Table 3.1) was then used as the basis for all superimpositions onto 1dlo. All superimpositions were performed in the InsightII® program,[152] as were RMSD and displacement value measurements. The displacement data was graphed using Excel®[153] (for further details see Chapter 6 Experimental).

The superimposition subset consisted predominantly of residues in the palm and connection subdomains of the p66 subunit, and the bulk of the p51 subunit, as illustrated in Figure 3.5. The subset contained a total of 376 residues; 41% of the total aligned residues (residues that were resolved in all the crystal structures) or 38% of all residues of the complete enzyme. Only 28% of the residues of the p66 were considered to remain static upon inhibitor or substrate binding. The fingers, thumb, and part of the

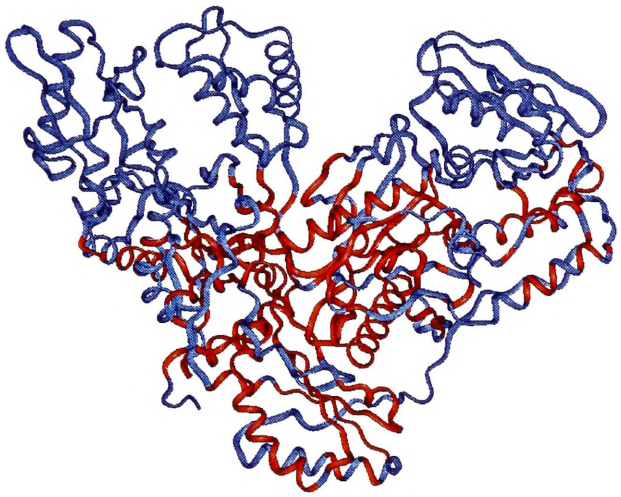


Figure 3.5 Ribbon diagram of the unliganded HIV-1 RT (1dlo) illustrating the distribution of the residues of the superimposition subset (red). The palm and connection subdomains of the p66, and most of the p51 subunit showed little conformational variation upon substrate or inhibitor binding.

RNase H subdomains demonstrated conformational changes greater than 2 Å. About 45% of the residues of the p51 subunit showed no conformational change when compared against the unliganded structure, supporting observations that the p51 acts as a scaffold for the catalytically active p66 subunit.

Although only the C_α atoms were used in the DDM calculations, the entire backbone of those residues deemed static was included in the superimposition subset. Considering the limited torsions and conformations possible of the backbone between two static C_α atoms, the inclusion of the backbone was deemed warranted and allowed for a justifiable increase in the number of atoms in the superimposition subset. This was verified by a comparison of the superimposition of the 376 C_α atoms versus the 1528 atoms of the backbone of those same residues, and by the RMSD values for the superimposition of all applicable structures onto the unliganded structure 1dlo (Table 3.2). A deviation in the RMSD values of less than 0.05 was noted for the two DDM derived methods, unquestionably attributable to the marked difference of the size of the datasets.[151]

PDB file	RMSD (Å)	Research Group
1hmv	0.878	Harvard I
1rtd	0.733	Harvard II
2hmi	0.997	CABM
1bqm	1.770	CABM
1bqn	1.809	CABM
1hni	1.807	CABM
1hnv	1.823	CABM
1klm	0.967	OCMS
1rev	1.204	OCMS
1rt1	1.035	OCMS
1rt2	0.975	OCMS
1rt3	1.173	OCMS
1rt4	1.081	OCMS
1rt5	1.082	OCMS
1rt6	1.052	OCMS
1rt7	1.073	OCMS
1rth	1.142	OCMS
1rti	1.033	OCMS
1tvr	1.829	CABM
1uwb	1.855	CABM
1vrt	1.304	OCMS
1vru	1.385	OCMS
1dlo		CABM

Table 3.2 The RMSD of RT crystal structures after superimposition onto unliganded RT (1dlo) using the DDM derived superimposition subset (1528 atoms). The different research groups, and their associates that resolved the crystal structures have also been noted; CABM (Center for Advanced Biotechnology and Medicine, Rutgers University, New Jersey), OCMS (Oxford Centre for Molecular Sciences, Oxford), Harvard I (Harrison research group, Department of Molecular and Cellular Biology, Harvard University, Cambridge), Harvard II (Verdine research group, Department of Chemistry and Chemical Biology, Harvard University, Cambridge).

All root mean square deviation (RMSD) values for the superimposition in Table 3.2 were less than 2 Å, with those of the unliganded (1hmv) and substrate-bound (2hmi and 1rtd) structures being less than 1 Å. The RMSD values also showed a preliminary trend related to the source of each crystal structure. Those crystal structures reported by the CABM (Center for Advanced Biotechnology and Medicine) research group had consistently higher RMSD values than structures reported by the OCMS (Oxford Centre for Molecular Sciences) group for the superimposition onto the unliganded structure, despite the fact that 1dlo was reported by the CABM group. The difference in the inhibitor-bound structures between the two research groups was on average 0.7 Å. This disparity between the groups of crystal structures was also evident in later aspects of the superimposition studies.

The resolution of the crystal structures used in our studies ranged between 2.2 Å and 3.8 Å, with an average of 2.8 Å. Although a difference distance of 2.0 Å was defined as being conformationally invariable, a value less than the lowest resolution of the studied structures, this was justified by the large number of crystal structures used and the consistency of results obtained in comparing the structures.

3.2.3 Comparison of superimposition techniques

The objectivity and validity of the DDM based superimposition was demonstrated by directly comparing the outcomes of superimposition using different reported techniques. In the first example, an unliganded (1dlo) and an inhibitor-bound (1vru) structure were superimposed onto a DNA-bound RT conformation (2hmi). The displacement of corresponding C_α atoms, in the region of the polymerase active site (residues 177-192), between the 2hmi structure, and 1dlo and 1vru were measured and plotted (Figure 3.6, the RMSD values and displacement data for these superimpositions are tabulated in Appendix 7.4).[139] The DDM based method used the 1528 backbone atoms of 376 static residues, while the Ding *et al.* technique used 16 C_α atoms of the β9-β10 hairpin (residues 177-192).

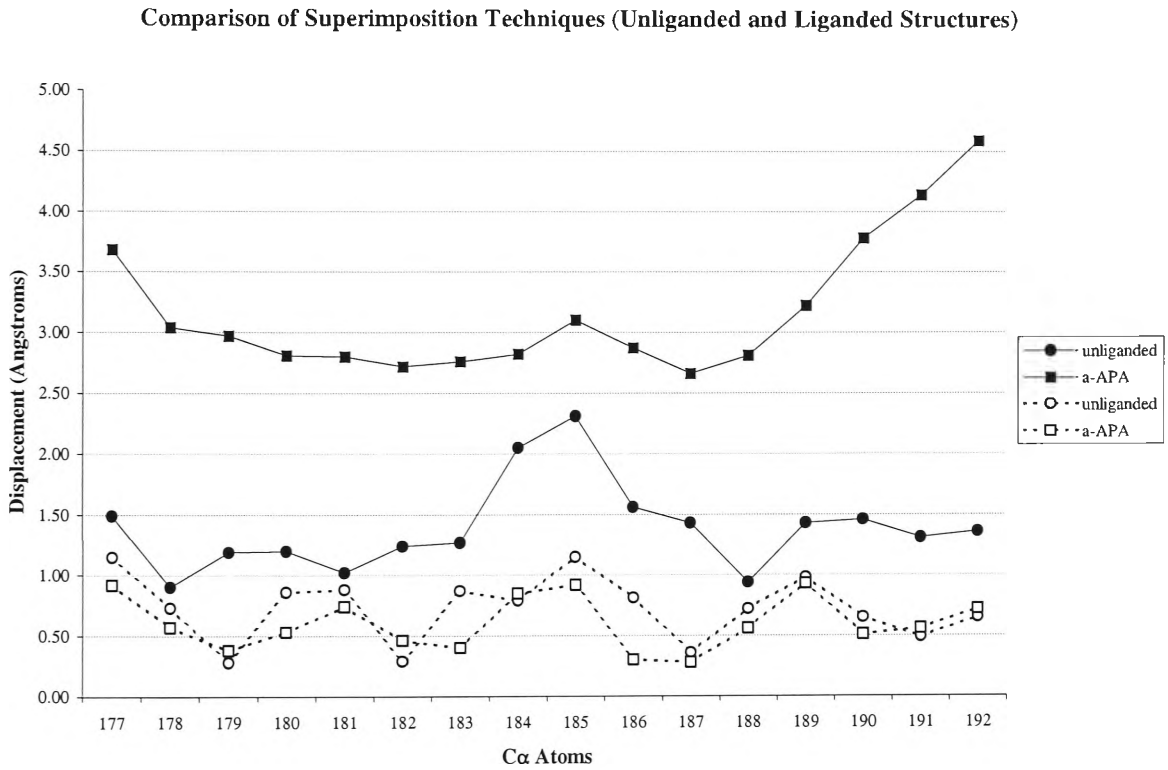


Figure 3.6 Displacement of Cα atoms of residues in the region of the polymerase active site upon superimposition of 1dlo (circle icons) and 1vru (square icons) onto 2hmi using different techniques. Solid icons represent the data from the DDM derived superimposition; while outlined icons represent the superimposition data using residues 177-192.

The RMSD values of the different superimpositions of the unliganded structure onto the DNA-bound structure differed by only 0.2 Å, while the inhibitor-bound structure superimpositions differed by a much greater 1.3 Å. As clearly seen in Figure 3.6, superimposition by the Ding *et al.* technique, that used the very residues to be

investigated, suppressed any conformational displacements that may have been induced by binding of DNA or α-APA. The average displacement of both structures using this technique was only 0.67 Å, with only two values greater than 1.0 Å. The DDM based superimposition conversely allowed residues demonstrating DNA or inhibitor induced conformational change to present themselves, despite the fact that some of these residues were also included in the superimposition subset (180-181 and 188-200). Most significantly were the polypeptide sections 184-186 and 188-192. Residues 184-186 form part of the catalytically significant polymerase active site. Using the Ding *et al.* technique, a slight propensity towards conformational change was seen in these residues, but was averaged out. The DDM based technique however allowed this propensity to be seen and impartially quantified. Also of significance were the residues 188-192, which showed no noticeable displacement trends using the Ding *et al.* superimposition. Even though these residues were included in the DDM derived subset, calculated from 19 different crystal structures, they demonstrated significant displacements, greater than the 2 Å defined as conformational invariability, in the inhibitor-bound structure which may merit further investigation.

In the second example, various inhibitor-bound RT structures were superimposed onto a RT structure complexed with nevirapine, using a published 'core' subset (data tabulated in Appendix 7.5).[138] The displacement of C_α atoms in the region of the non-nucleoside inhibitor binding site was measured and plotted (Figure 3.7). The average RMSD for the superimposition of four different inhibitor-bound crystal structures (1rth, 1vru, 1rti, and 1rev) onto the nevirapine-bound structure (1vrt) was 0.60, using the DDM derived subset, as compared to the 0.48 for the 'core' subset. This was remarkably comparable considering the average 'core' subset was composed of 104 atoms, 6.8 % of the number used in the DDM based superimposition.

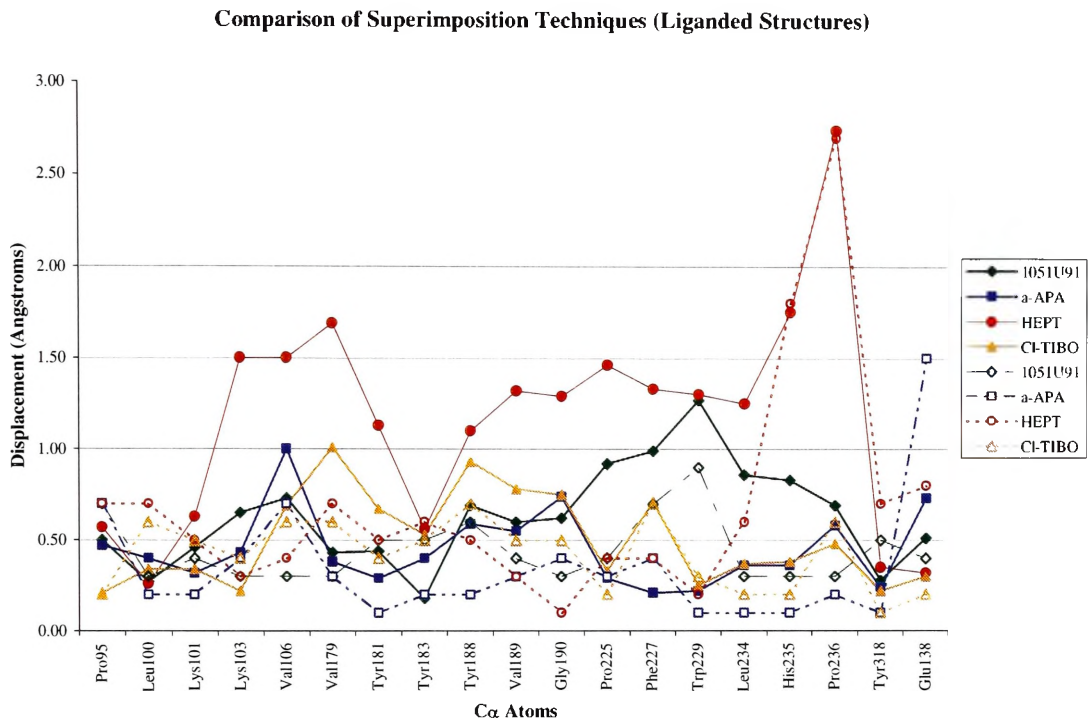


Figure 3.7 Displacement of Cα atoms of residues of the non-nucleoside inhibitor binding pocket upon superimposition of inhibitor-bound crystal structures onto 1vrt using different techniques. Solid icons represent the data from the DDM derived superimposition; while outlined icons represent the superimposition data using the 'core' subset.

Once again the DDM based superimposition allowed conformational differences between the inhibitor bound structures to be clearly seen and more accurately presented. The use of the 'core' subset was rational and in accord with the philosophy of the DDM based method, but its derivation lacked impartiality and autonomy from bias. The results from the 'core' superimposition showed a similar averaging effect to that seen in the first comparison of superimposition techniques.

The DDM based superimposition did not purposely magnify conformational displacements induced by substrate or inhibitor binding. Rather it allowed those conformational changes to be seen in their full magnitude without being masked by averaging out effects or suppressed by being forced to superimpose. Consequently it became a valuable and objective method to study and quantify ligand and substrate induced conformational changes at active sites and structurally important motifs in the RT enzyme. The circumspect interpretation of these studies was conducted and led to a greater understanding of the mechanisms of enzyme function and inhibition.

3.2.4 Superimposition for conformational studies

In the unliganded enzyme, the active sites and several structurally important motifs of the enzyme are found in a different conformation to that seen in substrate or inhibitor-bound structures. The structural changes required to adopt active or catalytic conformations, and those induced by the binding of non-nucleoside inhibitors, were studied in this work by the superimposition of all available unliganded, DNA-bound and inhibitor-bound crystal structures onto an unliganded structure (1dlo). The details of all crystal structures are given in Appendix 7.3. The displacement between the 1dlo structure and all superimposed structures was measured for each atom of the residues of the active site or structural motif studied.

1dlo was used as the unliganded structure because of its higher resolution and data completeness compared to the 1hmv structure, and because of irregularities in the crystal collection method of 1rtj, as detailed in Section 3.2.2. The second unliganded structure used (1hmv) provided further verification of the conformation of the primary unliganded structure, and substantiated the comparative displacements between unliganded, substrate-bound and inhibitor-bound structures; this structure was referred to when incongruities were observed between the unliganded structures. The two DNA bound structures were differentiated as the RT-DNA-dNTP catalytic ternary complex (1rtd) and the RT-DNA binary complex (2hmi). In these structures, the primer and template strands of nucleic acid are jointly considered as one DNA entity. To simplify the explanation and help visualise the displacements of residues, the structural analogy of the clenched right hand in Figure 3.3 was used to illustrate the conformational changes.

The crystal structures used in the superimposition studies (Appendix 7.3), were derived predominantly from two different research groups, the CABM and the OCMS research groups. In addition, two significant structures (1hmv and 1rtd) came from two separate groups at Harvard University, Cambridge. Most of the structures are in the *C2* or $P2_12_12_1$ space group, with 2hmi being the only exception in the $P3_212$ space group. Comparisons between space groups were made based on previous findings that differences in the crystallographic space groups did not significantly affect the structure of the enzyme.[147,151] Slight differential trends were noted between inhibitor-bound crystal structures that were obtained from one research group or the other, but these are small and in no way invalidate general inhibitor-induced structural changes noted in the

enzyme. These differential trends could be extended to the corresponding space groups as nearly all the structures of the CABM group are in the C2 space group, and the OCMS structures are in the $P2_12_12_1$ space group. The differences between the structures are most likely due to the process of structure optimisation and model refinement within each research group, rather than solely due to the different crystallographic space groups.

Superimposition using the objective DDM based technique provided greater certainty and insight into the conformational changes in the unliganded structure induced by substrate and inhibitor binding. Two active sites, that of the polymerase and RNase H domains, and two structural motifs that make contact with the nucleic acid hybrid, the primer and template grips, were studied in the following work to quantitatively demonstrate the differences in the position and conformation between structures. These differences were interpreted to provide insight into the activation and inhibition of the PAS. Additionally, residues of the inhibitor binding pocket were compared between unliganded, and substrate or inhibitor-bound structures to gain a greater appreciation of flexibility of the pocket and the structural requirements necessary for inhibitor binding.

3.3 The polymerase active site

3.3.1 The polymerase active site and polymerisation

The polymerase active site (PAS) of HIV-1 RT is composed of the highly conserved YMDD motif (Tyr183, Met184, Asp185, and Asp186) and Asp110 of the catalytically essential aspartic acid triad that includes Asp185 and Asp186. It lies on an antiparallel three-stranded β-sheet (β6-β9-β10) in the palm subdomain of the p66 subunit, and forms the principal biological machinery for the reverse transcription of viral RNA into proviral DNA.

The faithful insertion of nucleotides in DNA synthesis is believed to involve dynamic interactions between the enzyme, the nucleic acid and dNTP substrates, during which RT is proposed to undergo at least three conformational changes.[154] The first of these may occur when the nucleic acid primer-template binds to RT to form a tight E•DNA complex as shown in Figure 3.8.[123,155]

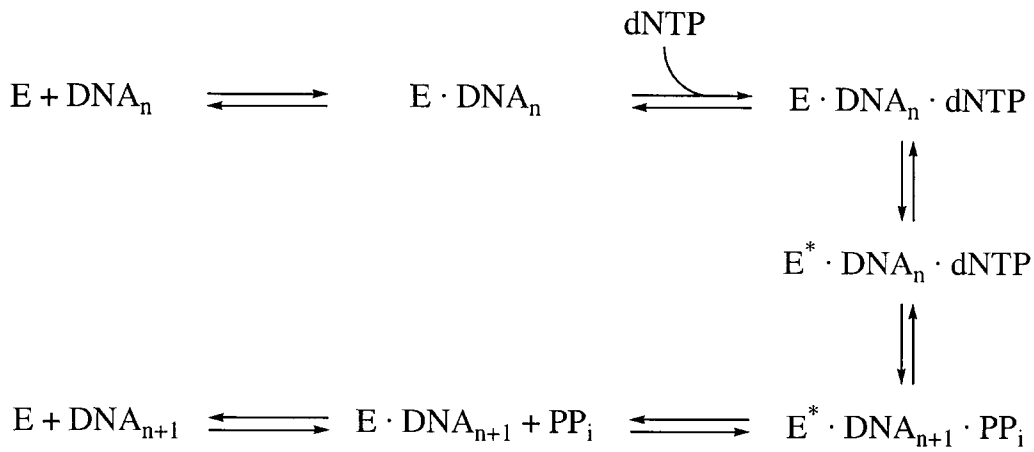


Figure 3.8 Catalytic pathway for polymerisation.[123] The abbreviations used: E, enzyme; DNA, deoxynucleic acid; n, nucleotide number; dNTP, deoxynucleotide triphosphate; E*, catalytically active enzyme, PP$_i$, pyrophosphate.

The binding of a dNTP is postulated to result in the formation of a loose initial ternary E•DNA•dNTP catalytic complex, where the primer and template strand of the DNA are considered as one entity. This complex is believed to undergo a rate-limiting conformational change to form a tight, enzymatically competent E*•DNA•dNTP complex, providing the dNTP has Watson-Crick complementarity to the appropriate nucleotide of the template. This putative conformational change is thought to facilitate the interaction of the incoming nucleotide with the 3'-OH of the primer, allowing a critical transition to be reached leading to the rapid chemistry of polymerisation forming the phosphodiester bond.[123,156,157] After nucleotide incorporation, a final conformational change could allow the enzyme to return to the loose state in order to release the pyrophosphate (PP$_i$), and translocate toward the new 3'-OH primer terminus to initiate the next round of reaction.[123] Overall HIV-1 RT can complete this catalytic cycle 20 times per second.[156,158]

The prevalent understanding that it is the rate-limiting conformational change that governs the maximum rate of incorporation of the correct nucleotide, and not the chemical step involving nucleoside phosphoryl transfer,[123] accentuates the significance of the structural changes needed for a catalytic conformation. This can be further appreciated with an understanding of the intricate interplay of the residues of the PAS, template-primer, dNTP and divalent metal ions involved in polymerisation.

The recent crystallisation of a 'catalytic' complex of RT with a template-primer and dNTP poised for catalysis, has corroborated hypotheses concerning the chemistry of dNTP polymerisation *via* an S_N2 displacement of pyrophosphate by the 3'-hydroxyl at the primer terminus.[136,159]

The most commonly accepted mechanism employs two Mg^{2+} ions, one catalytic (A) and one nucleotide binding (B), as depicted in Figure 3.9 from the catalytic 1rtd structure. These cations are thought to facilitate an S_N2 nucleophilic attack of the 3'-OH of the primer terminus on the α-phosphorus of the incoming dNTP and stabilise the negative charge of the transition state.[123,136,160]

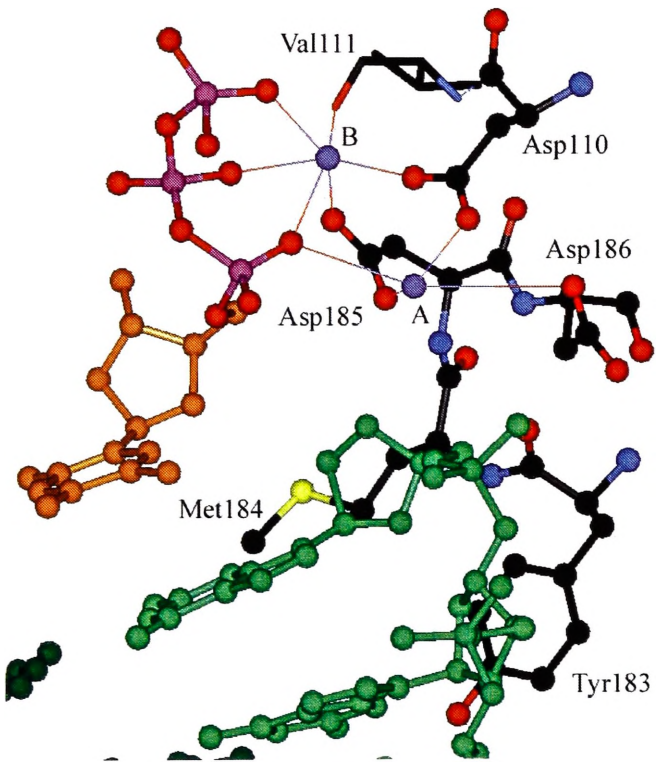


Figure 3.9 Proposed coordination complex of polymerisation in the catalytic ternary crystal structure (1rtd), where the thin lines to the Mg ions represent postulated interactions. The phosphate backbone of the dNTP and the residues of the PAS are coloured by atom, phosphorus (magenta) oxygen (red), carbon (black), nitrogen (blue), sulfur (yellow); the base of the dNTP (orange); Mg ions (purple), primer strand (light green), and template strand at the bottom left hand corner (dark green). Residue Val111, which does not form part of the PAS but is believed to be coordinated to the nucleotide binding Mg ion, is shown as sticks.

The catalytic Mg^{2+} ion (A) is speculated to contact the 3'-OH of the primer terminus, and its coordination shell most likely includes the sidechains of all three aspartic acid residues, an oxygen of the α-phosphate of the dNTP and probably a water molecule. The 3'-OH is actually missing from the primer terminus, as seen in Figure 3.9, to help capture the catalytic crystal structure at the stage just prior to polymerisation, and prevent actual bond formation between the primer strand and the incoming dNTP. Thus the 3'OH position and possible contact with the Mg ion can only be postulated. The triphosphate moiety of the dNTP is wrapped around the second Mg^{2+} ion (B), with each of the phosphates contributing a non-bridging oxygen to the octahedral coordination of

the Mg^{2+} ion. The contributions of the sidechains of Asp110 and Asp185 and the backbone carbonyl of residue 111 complete its coordination complex. This second Mg^{2+} ion likely contributes to the electropositive character of the α-phosphate facilitating a nucleophilic attack by the oxygen atom of the 3'-OH primer terminus. The elongated primer would be readied for the next polymerisation reaction and the by-product, PP_i, is thought to transiently bind to the aspartates via Mg^{2+} coordination prior to its release.[154] This explanation illustrates the numerous structural elements that may be required to converge in a precise and discriminatory mechanism for the successful orchestration of the polymerisation reaction.

Virtually all literature on the crystallographic studies of RT has focussed on the conformational changes of residues in the non-nucleoside inhibitor binding pocket for drug design and optimisation purposes. The few that refer to changes in the active site have made vague qualitative observations of differences between only a few enzyme structures; the comparison techniques have not been objective, results not quantitative, and have been without detailed analysis of specific residues. One study had postulated a mechanism of inhibition after comparing three inhibitor-bound crystal structures against an unliganded structure (1rtj).[29] Unfortunately, this unliganded structure was produced by soaking out a weak binding non-nucleoside inhibitor from pregrown crystals, making the observations from this structure dubious at best (see Section 3.2.2). Our studies have begun to rectify the lack of detail in this area of research by looking at the individual residues of the active site in all the crystal structures currently available.

3.3.2 Synopsis of the displacement of the PAS

The results of the superimposition showed clearly the differences in conformational change and displacement that are induced upon substrate or inhibitor binding, illustrated in Figure 3.10 and quantitatively summarised in Table 3.3 and Figure 3.11. The residues of the PAS are listed in Appendix 7.6, and displacement data is tabulated in Appendix 7.7. The data for Figure 3.11 was generated from superimpositions and distance measurements performed in the InsightII® program,[152] and graphed using Excel®[153] (for further details see Chapter 6 Experimental).

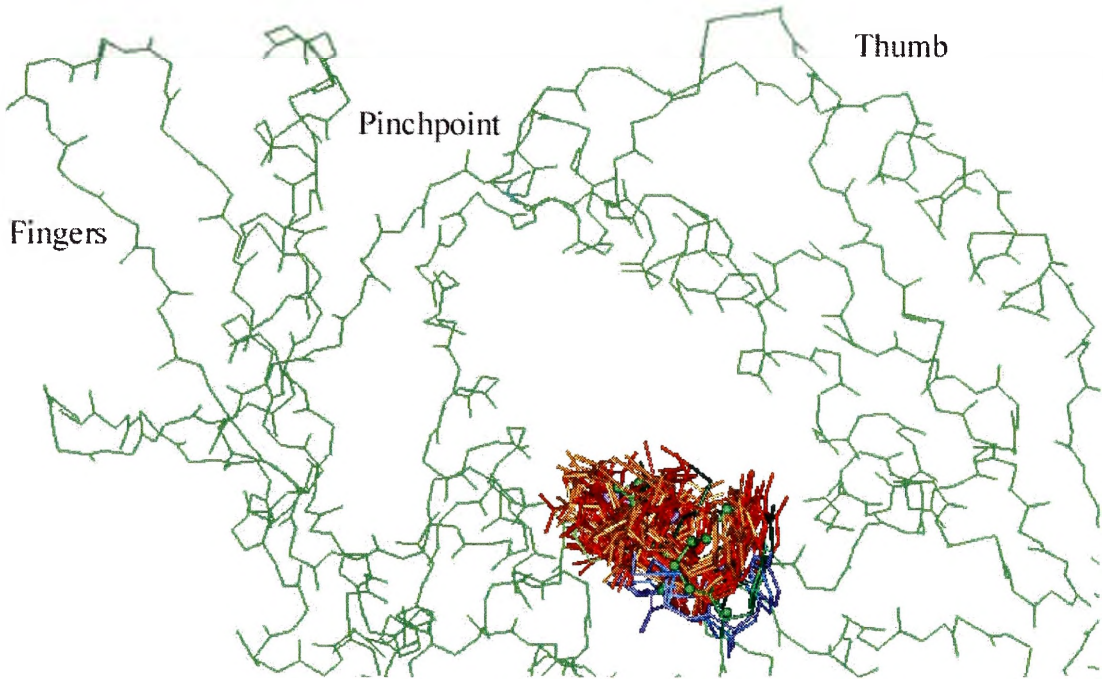


Figure 3.10 An illustration of the displacement of the PAS, in the palm subdomain, upon substrate and inhibitor binding relative to the fingers subdomain, on the left, and the thumb subdomain (in the 'closed' conformation), on the right. The backbone of 1dlo is shown as a green line trace. Red and orange active sites represent inhibitor-bound structures, light and dark blue active sites represent DNA-bound structures. This view is the same used in the individual residue analysis, with details of the conformational changes in the text.

Although the results in Figure 3.11 were only scalar, and did not express any direction of movement, they provided an initial snapshot description of overall conformational changes, showing the clustering of inhibitor-bound structures, and consistent changes of residues in a clear concise fashion. The breakdown of the average displacement into backbone and sidechain components (Table 3.3) indicated that the sidechains of all structures experienced additional conformational changes to those induced by the backbone. This additional conformational displacement of the sidechain was greater in the inhibitor-bound structures than the substrate-bound structures.

The comparison with the second unliganded structure (1hmv), on the whole, substantiated the comparison in conformational changes between all the studied structures and the unliganded structure (1dlo). The average displacement of the backbone of 1hmv was less than 1.0 Å, and the backbone and sidechain elements also showed much smaller displacements than those seen in the inhibitor-bound structures; in Figure 3.12, only 2 atoms showed a displacement greater than 2 Å. The binary RT-DNA (2hmi) structure showed similar conformational changes, in the backbone and sidechain, to the ternary (1rtd) structure, but these were all greater in magnitude. No explanation can be provided for the greater conformational similarity of 1rtd and 1dlo

compared to 2hmi and 1dlo. The 2hmi structure is better resolved than the 1rtd structure, 2.8 Å compared to 3.2 Å, although it has been refined from an earlier structure.[132,139] The inhibitor-bound structures all showed considerable displacements of the PAS, ranging from an average of 2.1 – 3.4 Å, with some individual atoms moving as much as 7.9 Å (Figure 3.11). More significantly the consistency and prevalent trends of the line traces in Figure 3.11 provided evidence that there were patterns of conformational changes in the PAS among the structures, and differences in the residues were not merely random and attributable to entropy. The crystal structures of the CABM research group had greater average differences in the conformation of the PAS (3.06 Å), compared to the unliganded structure, than those from the OCMS research group (2.54 Å).

average displacement (Å)		DNA-bound structures		inhibitor-bound structures		
	1hmv	ternary	binary	total	CABM	OCMS
entire residue	1.24	1.69	2.36	2.70	3.06	2.54
backbone	0.92	1.31	1.85	2.15	2.59	1.95
sidechains	1.52	2.01	2.80	3.18	3.47	3.05

Table 3.3 Summary of the average displacement between the 1dlo unliganded structure and the second unliganded (1hmv), substrate-bound (1rtd-ternary and 2hmi-binary), and inhibitor-bound structures for the entire PAS, and only the backbone or sidechain component. The three representative average displacements were calculated for all 19 inhibitor bound structures (total), and based on the research groups that derived the crystal structures; six from the CABM research group, and 13 from OCMS research group.

3.3.2.1 Substrate induced conformational changes of the PAS

The PAS of the binary and ternary dsDNA crystal structures were remarkably similar in position and conformation to each other, as can be seen in Figure 3.11 and more clearly in Figure 3.12. The line-traces for 2hmi (grey diamond) and 1rtd (black circle) followed similar trends and have comparable magnitudes of displacement. In Figure 3.13 (left), the thin stick structure of 2hmi is more similar to the pale structure of 1rtd, than the thick stick structure of 1dlo. DNA binding caused the PAS to be depressed into the palm comparative to the unliganded structure, illustrated in Figure 3.10 and more clearly seen in the later individual residue assessments (Section 3.3.3 – 3.3.7).

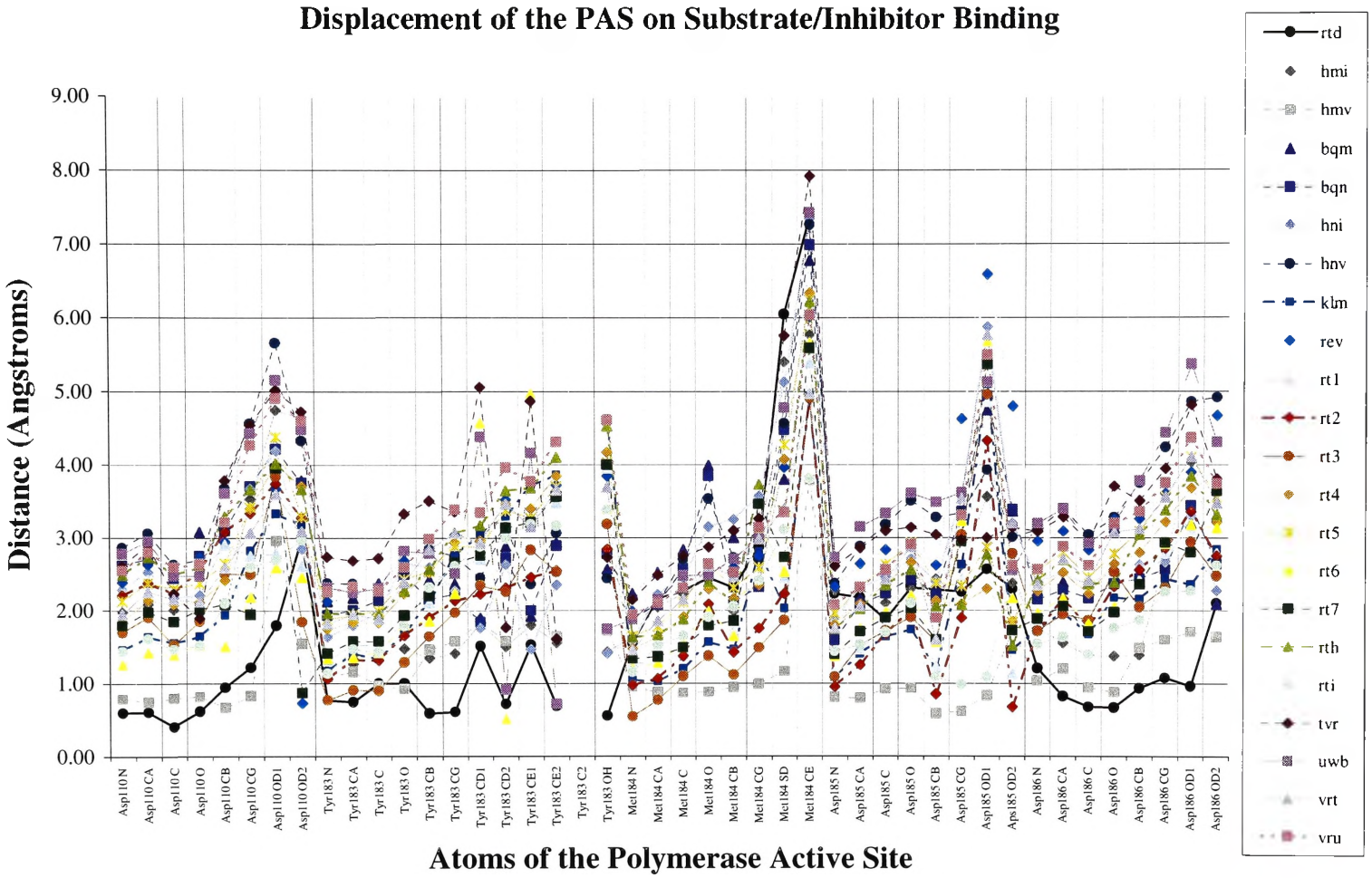


Figure 3.11 Graphical representation of the displacement between the unliganded (1dlo) structure and all other RT structures for the PAS residues.

The YMDD motif also showed a very small displacement forward away from the back of the hand, and very slight displacement left toward the fingers. The largest differences in displacement were seen in the sidechains of Asp110 (3.0 - 4.8 Å), Met184 (3.8 - 7.4 Å), and Asp185 and 186, with notably consistency. Figures 3.12 and 3.13 (left) illustrate the relative similarity between the unliganded and DNA-bound structures, with differences being predominantly seen in the sidechains of the PAS.

In Figure 3.13, the changes in conformation are seen relative to the catalytic 1rtd structure instead of the unliganded 1dlo. It is uncertain if these different conformations of the unliganded (1dlo and 1hmv) and 2hmi structures could still suitably coordinate the PAS and the Mg ions as in the 1rtd structure of Figure 3.9. These small differences may be sufficiently distinct from the catalytically active conformation such that these structures exist as different forms of the enzyme as postulated in Figure 3.8. These results provide evidence that the conformational changes required of the PAS to transform from an unliganded inactive form or a substrate-bound inactive form, as suggested by the 2hmi structure, to a catalytically active conformation such as that of E^*•DNA•dNTP from Figure 3.8, are subtle but distinct.

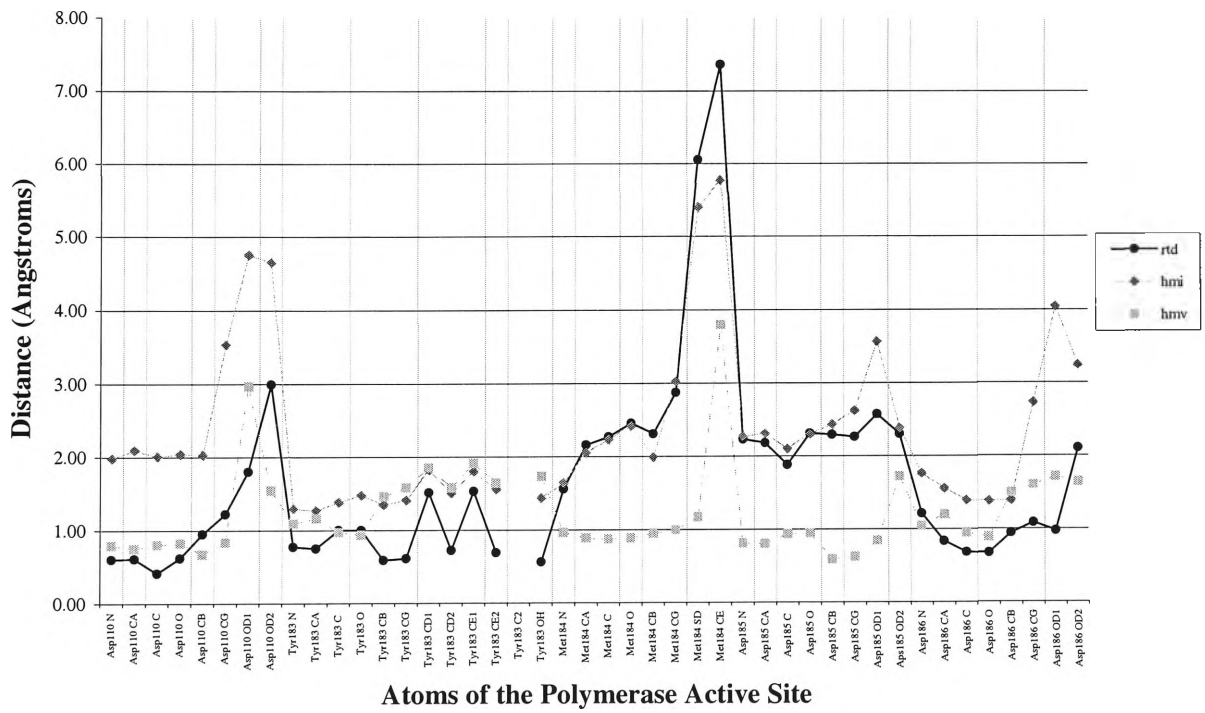


Figure 3.12 Graphical representation of the displacement between the unliganded (1dlo) structure, and the second unliganded (1hmv) and DNA-bound structures (1rtd and 2hmi) for the PAS.

3.3.2.2 Inhibitor induced conformational changes of the PAS

Inhibitor binding induced several concerted movements that displaced the PAS from its position and conformation in the unliganded state. The active sites of inhibitor-bound crystal structures showed consistent displacement left toward the fingers illustrated in Figure 3.10, as well as a small displacement toward the back of the hand. These displacements are quantitatively expressed in the multi-coloured line traces of the inhibitor-bound structures in Figure 3.11. Along with Figure 3.13, which illustrates the changes in conformation relative to 1rtd instead of 1dlo, this shows the basic and distinct differences in the magnitude of the conformational changes between inhibitor-bound structures and all other structures

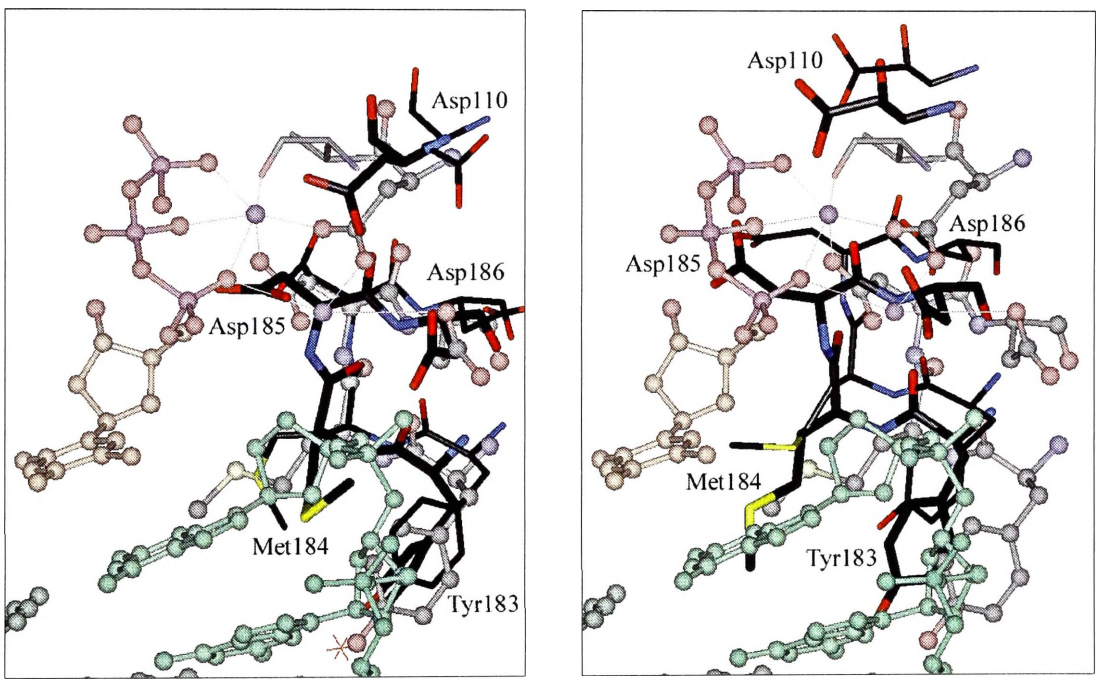


Figure 3.13 (Left) The conformational differences between the unliganded structure 1dlo (thick dark sticks) and the binary DNA structure 2hmi (thin dark sticks), seen relative to the proposed catalytic coordination complex of polymerisation in 1rtd shown in pale ball-and-stick (see Figure 3.9 for full legend details). (Right) Inhibitor-bound structures of 1hni (thick dark sticks) from the CABM group and 1vru (thin dark sticks) from the OCMS group, also seen relative to 1rtd.

The disparity between crystal structures derived from the different research groups was seen in the average displacement of the PAS in Table 3.3, where the CABM group structures were displaced on average by 3.06 Å, compared to 2.54 Å for the OCMS structures. This was also observed in Figure 3.13 (right), where the thick stick structure of 1hni (CABM) is displaced further to the left than the thin stick structure of 1vru (OCMS). Those structures from the CABM group tended to display a slight additional

elevation from the palm relative to the unliganded structure, not seen in these illustrations.

The comparison of the inhibitor-bound structures of Figure 3.13 (right) to the catalytic 1rtd shows clearly the active site residues being displaced to the left toward the fingers, intruding into the space occupied by the pale dNTP, primer strand and Mg ions in the catalytic structure. The coordination of the residues of the PAS, the Mg ions, as well as the incoming dNTP and the primer strand being would be difficult after such distinct conformational changes as observed in these studies.

3.3.3 Conformational changes for aspartic acid 110

The second unliganded structure (1hmv) showed a virtually identical conformation to that of the unliganded structure (1dlo). Only a negligible shifting toward the back of the hand accounted for the small displacement seen in the light grey square line trace of the backbone of this residue in Figure 3.12. The black line trace of the catalytic DNA complex (1rtd) also showed a similar orientation, but the residue was slightly forward of 1dlo, in the direction of the RNase H subdomain, as illustrated by the dark blue residue on the left in Figure 3.14. The binary DNA complex (2hmi) overlaid that of 1dlo, with a small downward displacement into the palm. From Figures 3.11 and 3.12 it can be seen that the grey diamond line trace for 2hmi is similar to the inhibitor-bound structures but, as seen in Figure 3.14, the direction of the displacement is distinctly downward into the palm, for 2hmi, rather than left toward the fingers subdomain. The Y-shaped carboxylate fork of this residue in the 2hmi structure was rotated to point toward the base of the thumb, in contrast to that of 1dlo that pointed toward the

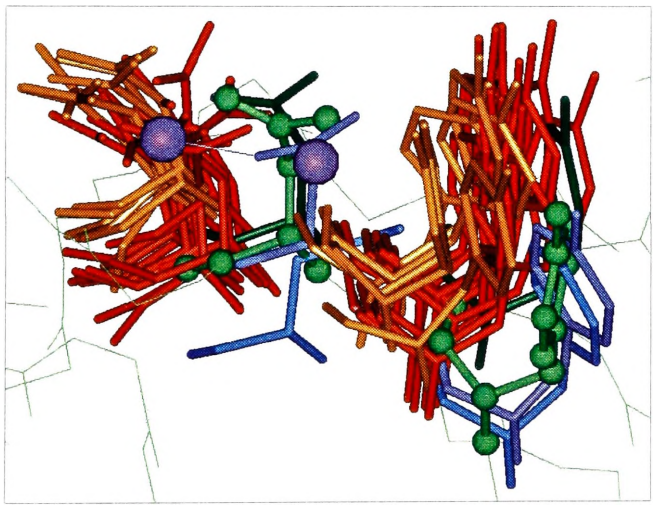


Figure 3.14 The displacement of Asp110 (residues on the left) and Tyr183 (residues on the right) upon substrate and inhibitor binding. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The backbone of 1dlo is shown as a green line trace, and the Mg ions in 1rtd are shown as purple spheres. The fingers subdomain is to the left of Asp 110 and the thumb subdomain is to the right of Tyr183, as can be more clearly seen in Figure 3.10, which is in the same orientation.

fingers and thumb pinchpoint (Figure 3.10).

The inhibitor-bound structures show distinct and uniform displacements toward the fingers, and smaller displacements toward the back of the hand. The coloured line traces in Figure 3.11 follow all the same trends, with displacements greater in magnitude than those of the second unliganded (1hmv) and catalytic ternary (1rtd) structures. The CABM structures, seen as orange residues in Figure 3.14, were clustered together and showed a slightly greater displacement further left toward the fingers than those of the OCMS group. Although vertical displacements were not of a serious magnitude, the CABM structures as a whole showed a very slight elevation compared to the OCMS group structures which were more on level with the unliganded structure. The conformations of the carboxylate forks were generally unchanged, with slight deviations pointing more toward the fingers than the pinchpoint. The clustering of the inhibitor-bound structures as a whole, and their difference from 1dlo and the DNA-bound structures, provided further support that small, inhibitor-induced conformation changes could be investigated in this manner, despite differences in crystal structure resolution and space groups.

Asp110 forms part of the catalytic triad of aspartates crucial to the function of the PAS, with any amino acid substitution dramatically impairing the catalytic activity of the enzyme. The catalytic rate of the polymerase reaction of a D110A mutant enzyme is 600 and 1000 fold lower than the wild type enzyme for RNA and DNA template-directed reactions respectively.[161] Interestingly, a serine substitution at this position selectively inactivated the RNA directed DNA synthesis but significantly retained the DNA directed catalysis.[161] Considering the contributions Asp110 makes to the coordination of both Mg^{2+} ions (Figure 3.9) we postulate that even a minor conformational change could leave the aspartate unable to participate in the stabilisation and activation of the metal ions. Its displacement could also be propagated along the polypeptide to residue 111, whose carbonyl backbone in Figure 3.9 was also postulated to make a contribution to the nucleotide-binding Mg^{2+} ion.[136] The positional differences seen between the two blue DNA-bound Asp110 residues in Figure 3.14 may signal a vital conformational change necessary to transform between the inactive DNA-bound form of the enzyme, to a form that is poised for polymerisation. We hypothesise that the different conformational changes seen in the orange and red inhibitor-bound structures of Figure 3.14 and the dark residues of Figure 3.13 (right) undermine the

ability of Asp110 to coordinate the nucleotide binding Mg ion necessary for catalysis. The binding of non-nucleoside inhibitors may 'lock out' the residue from more flexible conformations that could transform into passive and active or catalytic forms capable of coordinating the necessary cations and its surrounding residues.

3.3.4 Conformational changes for tyrosine 183

The tyrosine 183 of the 1hmv unliganded structure was slightly different from that of 1dlo, being positioned marginally more toward the back of the hand and slightly elevated from the palm. On the right hand side of Figure 3.14, the dark green residue of Tyr183 of 1hmv is almost hidden by the displacements of the inhibitor-bound structures, and the elevation is slightly exaggerated by the angle of the illustration. The phenyl sidechain in 1dlo was in an upright diagonal position, with the phenol group directed to a point above the palm/base of palm, in the direction of the nucleic acid groove. The phenyl ring, which faced the connection subdomain in the unliganded structure, was rotated about 40-50° to face the thumb with a very slight displacement down into the palm in 1hmv. The DNA-bound structures (1rtd and 2hmi), seen as dark and light blue sticks in Figure 3.14, were extremely similar to each other in position, as well as to 1dlo; with a similar rotation to that of the phenyl ring in 1hmv. The only displacement the DNA-bound structures showed was a slight displacement down into the palm; more noticeable in the 2hmi structure than 1rtd. These displacements were all less than 2 Å, and as such, could be considered as being conformationally invariable compared to the unliganded structure. The displacements followed a fairly consistent trend between the compared structures, as seen in the three line traces in Figure 3.12 for this residue.

Figure 3.14 illustrates how the inhibitor-bound crystal structures, as a group, showed a small displacement toward the back of the hand and a marginal elevation up from the palm compared to the unliganded 1dlo. The magnitude of these displacements in Figure 3.11 for Tyr183 was on average about 2.5 Å, compared to about 1.3 Å for the unliganded and DNA-bound structures seen separately in Figure 3.12. The CABM structures, seen as orange residues in Figure 3.14, showed a marginally greater upward displacement from the palm and a small displacement toward the fingers and less so toward the back of the hand than those structures from the OCMS group, which had negligible sideways displacement. The phenyl sidechains of the inhibitor-bound

structures were similarly swivelled as in 1rtd and 2hmi; in the OCMS group this was seen more consistently and clustered, but in the CABM structures this varied widely from 20-90° as illustrated in Figure 3.14.

Ty183 plays no role in the coordination of the divalent Mg ions, nor does it interact with the incoming dNTP during polymerisation. Rather, it is one of the few amino acids of RT that is thought to interact with bases in the minor groove of the template-primer at the PAS,[136] and its mutation can have varying effects on polymerase activity, processivity, and fidelity. Its lack of participation in the coordination complexes involving the divalent Mg ions supports the observations that mutations of Tyr183 show a reduction in polymerase activity, but never a total cessation of DNA incorporation.[162] Results of several other research groups have suggested that the hydroxy group and phenyl ring of the tyrosine appear to be crucial for the synthesis of the negative strand DNA while neither of these may have a significant role in the plus strand DNA synthesis.[163]

Tyr183 is thought to also play a role in maintaining enzyme processivity, which is a measure of the number of nucleotides incorporated into a primer strand during a single binding event before the enzyme dissociates from the template-primer. Substitutions of Tyr183 that increase the rigidity of an otherwise flexible dNTP-binding pocket are known to enhance the fidelity of polymerisation by accommodating only the correct nucleotide for catalysis. The mechanism for this occurs either during the initial binding step or by inhibition of the conformational change step after binding of the incorrect nucleotide.[162,163] The ability of Tyr183, unlike its mutants, to both incorporate incorrect nucleotides and extend mispairs on RNA templates,[163] as well as its influence in enzyme processivity, alludes to a role in an evolutionary strategy that introduces variances in the genome which may be beneficial to the virus. This is significant mechanistically, and in the consideration of drug design since DNA synthesis on the RNA template is the more vital of the two rounds of synthesis; subsequently the RNase H digests the RNA template, and this primary copy of the viral genome is lost.

The sidechain of Tyr183 can be seen, in Figure 3.9, to be quite close to the backbone of the penultimate nucleotide of the primer strand, possibly participating in its stabilisation. The DDM-based superimposition and comparison of unliganded and substrate-bound enzymes in Figure 3.13 (left), as well as the line traces of this residue in Figure 3.12, showed little differences in conformation. Any variance could be passed

off as natural flexibility in this residue of the dNTP-binding pocket, to accommodate a nucleotide, or different nucleotides. In contrast, the inhibitor-bound structures in Figure 3.13 (right) showed more notable differences in conformation to the unliganded structure, with Tyr183 displaced further toward the last nucleotide of the primer strand, that is poised to undergo polymerisation with the incoming nucleotide. We postulate that the displacement of Tyr183 in the inhibitor-bound structures disrupts the positioning and 'reading' of the last nucleotide of the primer strand, leaving it in a position that is not optimal for the incoming nucleotide to interact with. Furthermore, the conformational changes to Tyr183 may be disseminated to other neighbouring residues that in turn may play different and significant catalytic and structural roles in polymerisation.

3.3.5 Conformational changes for methionine 184

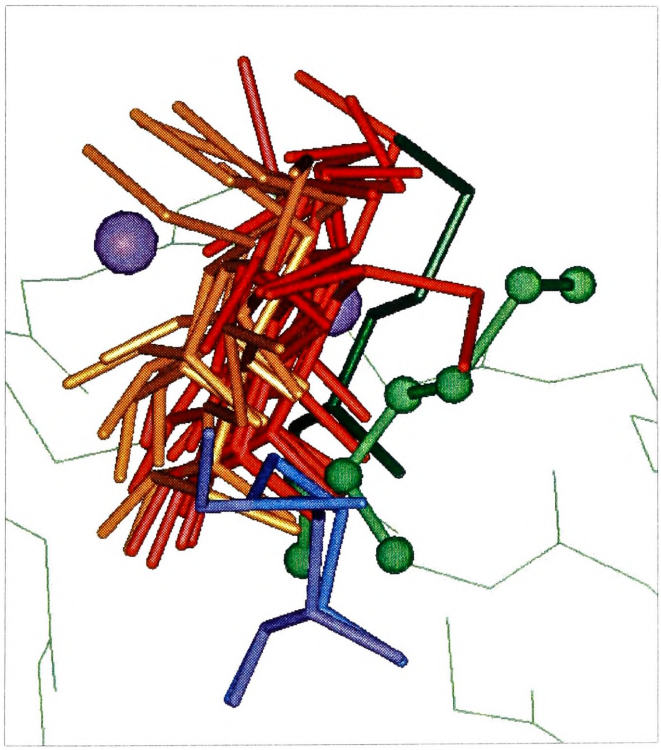


Figure 3.15 The displacement of Met184 upon substrate and inhibitor binding. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The backbone of 1dlo is shown as a green line trace, and the Mg ions in 1rtd are shown as purple spheres. The fingers subdomain is to the left of Asp 110 and the thumb subdomain is to the right of Tyr183, as can be more clearly seen in Figure 3.10, which is in the same orientation.

The alkyl sidechain of this residue in the unliganded structure pointed upward toward the tip of the thumb, with the last atom of the sidechain pointing toward the base of the thumb, in a direction toward the back of the hand. The second unliganded structure 1hmv, by comparison, showed a small displacement toward the back of the hand; retaining the same direction in the sidechain but with the last atom of the sidechain pointed toward the fingers, seen as the dark green residue in Figure 3.15. The DNA-bound structures both showed a downward displacement into the palm,

clearly illustrated by the light and dark blue residues in Figure 3.15, and a slight displacement forward, away from the back of the hand. Their sidechains angled diagonally upward across the palm toward the smallest finger, except for the last atoms of the sidechain which for the 2hmi structure pointed toward the RNase H and for 1rtd pointed to the fingers/thumb pinchpoint. The last two atoms of Met184 (S_δ and C_ε) caused the greatest displacement, seen in Figure 3.12, for the unliganded and substrate-bound structures. These displacements were of a similar magnitude to that seen in the multi-coloured line traces of the inhibitor-bound structures of Figure 3.11 but, as Figure 3.15 illustrates, were in quite a different conformation and direction.

The inhibitor-bound structures seen in Figure 3.15 showed a displacement left toward the fingers subdomain and a marginal displacement toward the back of the palm. The orange CABM structures showed a very slight vertical elevation, and tended to be displaced furthest to the left of all the structures. This vertical displacement can be seen in the backbone of CABM structures, at the base of the residues in Figure 3.15. Practically all of the sidechains of the inhibitor-bound structures were similarly vertical as 1dlo, with the last atom of the sidechain consistently pointing toward the fingers, accounting for the greatest displacements seen throughout the entire PAS.

Met184 is the X-residue in the catalytically important YXDD motif, which is highly conserved in retroviral RTs.[164] It lies in close proximity to the 3'-OH primer terminus and the bound dNTP, as seen in Figure 3.9, but plays no part in their coordination with either Mg^{2+}ion. The methyl group of the sidechain, C_ε in Figures 3.11 and 3.12, is believed to interact with the sugar moiety of the substrate dNTP, as do the C_α and C_β carbons with the ribose ring of the primer terminus.[139,164] Figure 3.13 (left) shows small differences between the unliganded and the substrate-bound structures of Met184 that may be due to the presence of the primer strand or the nature of the last nucleotide of the primer strand. The significantly greater inhibitor-induced conformational changes seen on the right of Figure 3.13 very likely disrupt the positioning of the 3'-terminus and interfere with the configuration it adopts with the incoming nucleotide to allow the chemical reaction of polymerisation to occur.

Met184 plays similar roles to that of Tyr183 in the dNTP-binding pocket, with effects on polymerase activity and fidelity. Changes in the size and bulk of the 184 sidechain appear to be factors in determining the catalytic activity of the enzyme; long and bulky sidechains seem to be detrimental to the polymerase activity, whereas the presence of

smaller side chains exhibits no inhibitory effect on the activity.[164] The specificity of methionine in this position is evident in the experimental data that shows that the difference between an alanine and valine can produce a highly error-prone enzyme or one exhibiting increased fidelity, without compromise to the efficiency of DNA synthesis.[164,165] Similar to Tyr183, the role Met184 plays in the efficiency and fidelity of DNA synthesis is conditional on the nucleic acid template, suggesting that this sidechain may be dispensable in terms of enzymatic activity, but is essential for nucleotide selection during DNA-directed second strand (+ strand) DNA synthesis.[165]

The displacements of the methionine sidechain in inhibitor-bound structures seen by our DDM-based superimposition, and illustrated in Figure 3.13, could make it unable to correctly align the 3'-OH of the primer terminus with the α-phosphate moiety of the dNTP. This could only be postulated, as the 3'-OH of the primer terminus is not present in the catalytic 1rtd structure.

PDB file	ϕ (°)	ψ (°)
1dlo	60.51	-160.58
1hmv	70.43	-147.55
1rtd	58.33	-106.33
2hmi	54.27	-106.86
1bqm	57.64	-83.76
1bqn	44.77	-94.53
1hni	14.66	-86.75
1hnv	-	-100.12
1klm	55.21	-130.89
1rev	54.50	-136.63
1rt1	55.03	-122.74
1rt2	54.11	-108.50
1rt3	49.03	-138.82
1rt4	62.45	-120.72
1rt5	60.45	-118.01
1rt6	57.04	-131.99
1rt7	57.08	-131.56
1rth	68.18	-123.16
1rti	57.56	-117.00
1tvr	51.2	-116.51
1uwb	56.92	-141.02
1vrt	53.80	-103.84
1vru	62.61	-116.35

Table 3.4 ϕ and ψ angles of Met184. ϕ (N, C_α, C, N_{+1}) and ψ (C_{-1}, C, C_α, C).

It has been suggested that the conformational changes in its backbone may also affect the stabilisation of the dNTP-binding pocket by affecting the β9-β10 hairpin. This hairpin was observed to have an unusual β-turn geometry, characterised by torsion angles of ϕ ~ 60° and ψ ~ -160° for Met184 in the unliganded enzyme.[29,135,139] Upon both substrate and inhibitor binding, Ding *et al.* observed that the ϕ angle remained relatively unchanged while the ψ angle increased to around -120°, which is the ideal value for a type II' β-turn. However, measuring the ϕ and ψ angles for Met184 in each of the structures used in our studies (Table 3.4) showed the ψ angles for 1rtd and 2hmi were virtually identical at -106°. They differed from the unliganded structure by about 50° (40° in comparison to 1hmv), supporting the notion of a substrate-induced conformational change. The ψ angles of

the inhibitor-bound structures showed insufficient consistency, ranging from -83° to -141°, for any definitive correlation between the structure of the β-turn and inhibitor binding to be made. This reiterates the need for any study of RT to include as many different structures as possible, so that a more complete and comprehensive account can be made of the observations. Conformational changes in Met184 were expected to propagate throughout the local region because of the stabilising hydrogen bonds methionine made with surrounding residues, effecting the stability of the dNTP binding pocket and thus dNTP substrate binding and polymerisation.

3.3.6 Conformational changes for aspartic acid 185

The orientation of the 1hmv structure was similar to 1dlo, as demonstrated by the light grey line trace in Figure 3.12, but showed a small displacement toward the back of the hand and a slight displacement toward the thumb (Figure 3.16). The carboxylate fork of Asp185, which in 1dlo pointed upward toward the fingers and thumb pinchpoint, instead was directed more toward the thumb. The light and dark blue residues of the DNA-bound structures were virtually identical to each other in position and orientation in Figure 3.16. In comparison to 1dlo, they showed a definitive downward displacement into the palm, and a very marginal displacement forward along the nucleic acid groove and left toward the fingers. Their carboxylate forks pointed upward similarly to 1dlo, with differences in the torsion about the Cγ atom.

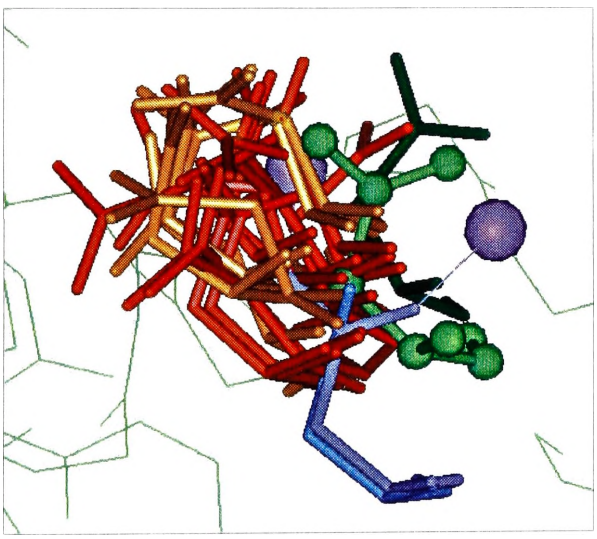


Figure 3.16 The displacement of Asp185 upon substrate and inhibitor binding. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The backbone of 1dlo is shown as a green line trace, and the Mg ions in 1rtd are shown as purple spheres. The fingers subdomain is to the left of Asp 110 and the thumb subdomain is to the right of Tyr183, as can be more clearly seen in Figure 3.10, which is in the same orientation.

A small displacement left toward the fingers and a slight displacement toward the back of the hand were the only discernible displacements seen in the inhibitor-bound structures. Most of the carboxylate forks of the inhibitor-bound structures pointed more horizontally toward the fingers, though a few maintained the same

direction toward the fingers and thumb pinchpoint as 1dlo. Asp185 in Figure 3.16 did not show consistent clustering of inhibitor-bound structures, as other residues have (Figure 3.14 and Figure 3.15), but the displacement trends in the coloured line traces of Figure 3.11 affirm quantitatively consistent inhibitor-induced conformational changes in this residue.

Asp185 is the second of the three essential aspartic acid residues of the PAS, its importance could be no more clearly illustrated than in Figure 3.9, where each of the oxygens of its carboxylate is involved in the coordination of the divalent metal ions. Experimental data from other research groups appears to indicate that this residue participates only in the forward reaction of polymerisation, by initially binding the incoming dNTP.[161,163,166] After the proposed conformational change to the enzymatically competent E^*•DNA•dNTP complex, and the initiation of the bond formation and cleavage process, Asp185 is believed to lose its coordination with the Mg ion and not participate in the reverse reaction of pyrophosphorolysis. Rather, the aspartate may interact directly with the 3'-OH of the primer, acting as a proton acceptor in a general base mechanism.[161]

The conformational changes of Asp185 induced by substrate binding are clearly seen in Figure 3.16, and were hypothesised to be necessary to accommodate the coordination complex of both Mg ions and the incoming nucleotide. The changes induced by inhibitor binding were on average 0.7 Å greater than substrate-induced changes. The average displacement of the multi-coloured line traces for the inhibitor-bound structures of Figure 3.11 was 2.6 Å and, as seen Figure 3.13 (right), left no doubt that the correct complexation of both Mg^{2+} ions, the incoming nucleotide and the template and primer strands would be near impossible to coordinate. The entire residue occupied the space where the nucleotide binding Mg^{2+} ion was positioned in the catalytic 1rtd structure, and began to encroach on the space occupied by the phosphates of the incoming nucleotide. It was also displaced too far toward the fingers domain to coordinate the catalytic Mg ion (Figure 3.16). It would be uncertain how much of an effect this would have on the ability of an incoming nucleotide to approach the 3'-terminus. Nevertheless we hypothesise that the conformational changes seen in Figure 3.13 (right) and Figure 3.16 would make it impossible to coordinate even the nucleotide, the Mg ions and the residues of the PAS, without considering the primer strand. Not only had Asp185 been repositioned, but its conformational changes were also propagated to its equally

important neighbouring residues. The displacement of Asp185 may not only interfere with its function in the polymerisation reaction, but may also affect other surrounding residues that may play roles in both forward and backward reactions that are equally important in the progressive DNA synthesis.

3.3.7 Conformational changes for aspartic acid 186

Only a small displacement toward the back of the hand and a slight displacement toward the fingers differentiated the second unliganded structure (1hmv) from 1dlo. The two green unliganded structures in Figure 3.17 were very similar, just as the line trace in Figure 3.12 remains below 2 Å. The DNA-bound structures show a small displacement downward into the palm, and a very slight displacement left toward the fingers and forward along the nucleic acid groove. The downward displacement is slightly greater in the case of the light blue structure of 2hmi (Figure 3.17), and the forward displacement is slightly more marginal in 1rtd. The carboxylate forks of these structures point toward the middle inside edge of the thumb for 1rtd and the base inside edge of the thumb for 2hmi. This is compared to 1dlo's carboxylate fork that points directly upward toward the tip of the thumb, with the plane parallel to the face of the thumb.

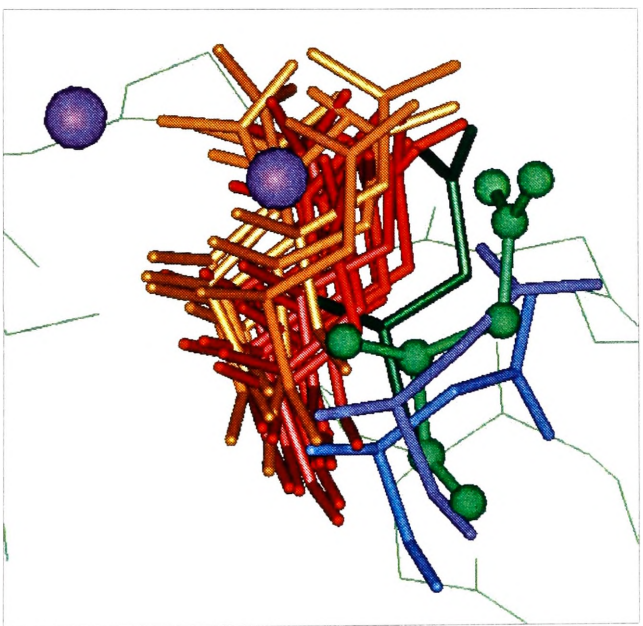


Figure 3.17 The displacement of Asp186 upon substrate and inhibitor binding. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The backbone of 1dlo is shown as a green line trace, and the Mg ions in 1rtd are shown as purple spheres. The fingers subdomain is to the left of Asp 110 and the thumb subdomain is to the right of Tyr183, as can be more clearly seen in Figure 3.10, which is in the same orientation.

The inhibitor-bound crystal structures showed a small uniform displacement left toward the fingers, and a slight displacement toward the back of the hand (Figure 3.17). Only a marginal elevation up from the palm was seen. The sidechains showed little conformational change or orientation, other than varying degrees of swivel about the Cγ atom.

The last of the crucial aspartic acid triad, the carboxylate of Asp186 coordinates the catalytic Mg^{2+} ion (A). Asp186 is known to be involved in both the forward and the reverse reactions[161] of DNA synthesis, most probably because of the significant position is takes in coordinating the catalytic Mg^{2+} ion. The changes in this residue from the unliganded position to the substrate-bound conformations, were postulated to be necessary to coordinate this cation in an optimal position for polymerisation. Asp186's precise coordination and geometry with this same ion is totally reordered by the inhibitor-induced conformational changes seen by the DDM-based superimposition, and illustrated in Figure 3.13 (right). These changes would most likely leave the carboxylate fork unable to coordinate to the Mg^{2+} ion, and also affect the positioning of the neighbouring Asp185 as well as the architecture of the β9-β10 loop to some degree.

3.3.8 Activation and inhibition of the PAS

The dynamic nature and function of RT demands that it be more than just a static structural protein. Disregarding possible coupled functions and communications it may have with other active sites and structural motifs elsewhere in the enzyme, the PAS is responsible for the DNA synthesis on both RNA and DNA templates. In a series of processes that are measured in microseconds, the PAS must coordinate the direct positioning of no less than five catalytic residues with two Mg^{2+} ions, nucleotides from each of the primer and template strands, as well as the incoming nucleotide. It must recognise and differentiate between ribonucleic and deoxyribonucleic acids, and correct and incorrect nucleotides. Its residues must function to find a balance between the necessity of flexibility to function and allow viral diversity, the hallmark of retroviruses, and rigidity to maintain genomic integrity and ensure viable progeny. Once all of these components are in place, conformational changes ensue which lead to the chemistry of polymerisation, translocation and the repositioning of all those elements of DNA synthesis once again.

The results from our DDM-based superimpositions showed that the unliganded RT enzyme was indubitably unable to bind DNA without conformational rearrangement. On a gross scale, the thumb subdomain needed to extend away from the fingers to accommodate the nucleic acid in the active site region. On a much more fine scale, the residues of the PAS needed to settle down slightly deeper into the palm region to accommodate the separate primer and template strands. The conformation of the PAS

in this E•DNA complex of the kinetic pathway (see Figure 3.8) was most likely represented by the 2hmi structure. We hypothesise that another conformational change of the PAS is required to incorporate the nucleotide into the E•DNA•dNTP complex before the rate limiting conformational change preceding nucleotide incorporation. That E^*•DNA•dNTP complex in the kinetic pathway was most probably represented by the catalytic 1rtd structure, and the conformational changes necessary to transform to this state were seen in the differences between the 2hmi and 1rtd structures. The differences in the PAS conformations of these DNA-bound structures to those in the unliganded structure were clear and irrefutable. The differences between the two DNA-bound structures were less glaring, but still notable, especially in the displacements of the three aspartic acid residues. This was perhaps to be expected considering their dominant role in coordinating the Mg^{2+} ions and the triphosphate moiety. A recent qualitative graphical comparison of the binary and ternary DNA-bound structures had found that there were rearrangements of the carboxylate sidechains of the PAS.[160] These rearrangements brought Asp110 close enough to chelate with the nucleotide binding Mg^{2+} ion (B),[160] but this published study did not provide further details or elaborate on the exact conformational changes of the remaining residues of the PAS.

Although a previous comparison of three different crystal forms of RT had proposed a rotation of the PAS relative to the rest of the enzyme,[143] no other study has been able to quantify and so clearly illustrate the specific flexibility needed to allow nucleic acid binding and polymerisation. This molecular swiveling, around an axis that is nearly parallel to DNA when bound to the large active site cleft, is attributed to an allowance for the rotational and translational movements of the growing nucleic acid duplex needed while the asymmetric $tRNA^{Lys3}$ acts as primer for first strand synthesis. This hypothesis can not be corroborated with our DDM-based conformational studies until RT crystal structures with RNA templates and primers become available.

Based on the premise that precise conformational changes are required at nearly every stage of the polymerisation process, it was anticipated that the failure of any one of these conformational changes to transpire, or the occurrence of an unforeseen conformational change, would disrupt the entire process. These DDM-based superimpositions have shown startling conformational changes that not only displace the PAS from its customary position, but may also immure it such that it cannot proceed to the next conformation necessary for polymerisation. Binding of non-nucleoside

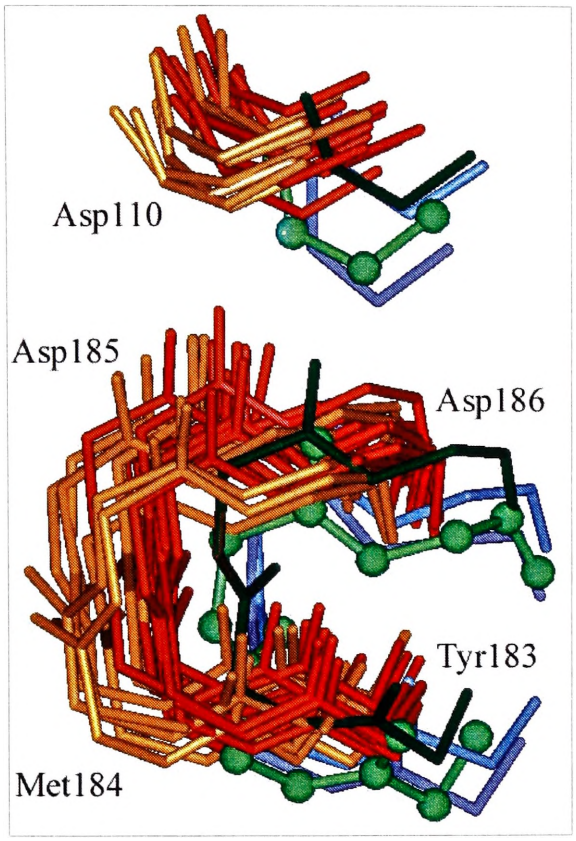


Figure 3.18 The displacement of the backbone of the PAS upon substrate and inhibitor binding. 1dlo (green ball-and-stick), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red).

inhibitors displaced the PAS predominantly toward the fingers subdomain, and slightly toward the back of the hand compared to both the unliganded and DNA-bound structures, as illustrated by the backbone of the PAS in Figure 3.18. Both DNA-bound structures, 2hmi and 1rtd, showed a displacement down into the palm away from the template-primer, supporting previous observations made only with the 2hmi structure.[139] A vertical displacement into the palm might have been expected in light of evidence that inhibitor-bound RT enzymes can bind various template/primers and nucleotides as well as, if not slightly better than, the unliganded enzyme.[167] However this may depend on how studies are conducted; with the inhibitor added to an E•DNA complex, as for the Spence *et al.* study, or with the primer-template added to an E•I complex, which would have been preferable for our studies. It is not presently known if a template-primer complex is able to bind to an inhibitor-bound enzyme with such conformational changes as identified by these superimpositions. The conformational changes we have studied in the available crystal structures of RT-inhibitor complexes were not affected by the presence of DNA, and may very well undergo an additional vertical displacement, or completely different conformational displacements, were DNA to be also present. No E•I•DNA crystal structures are presently available to confirm this speculation.

The rate-limiting step in the synthesis of DNA is a conformational change that is induced by the binding and recognition of the correct dNTP. Kinetically, the NNRTIs work by slowing the rate of the chemical reaction such that the conformational change is no longer rate-limiting,[123] but rather the chemical step is inhibited. Structurally, this may be explained by conformational rearrangements upon inhibitor binding that are less

than ideal for the polymerase machinery of the PAS, Mg^{2+} ions, template-primer and dNTP, which we have detailed in this study, slowing down the chemical reaction of polymerisation.

3.4 The primer grip

3.4.1 The primer grip and its role in reverse transcription

The 'primer grip' of HIV-1 RT consists of the β12-β13 hairpin (Phe227-His235) of the antiparallel three-stranded β-sheet (β12-β13-β14) in the p66 palm subdomain, which contacts the primer strand near its 3' terminus (Figure 3.19).[132,168] Highly conserved among retroviral RTs, the primer grip plays a role in aligning the 3'-OH end of the primer strand in the optimal position for nucleophilic attack on an incoming dNTP. Alanine-scanning mutations of primer grip residues has previously shown disparate impairment of both DNA-dependent and RNA-dependent DNA polymerase activities, as well as RNase H activity.[169,170-172] None of the known mutations that confer resistance to NNRTIs are mediated by changes in the amino acids of the primer grip, suggesting that these residues also have a critical architectural role.[172]

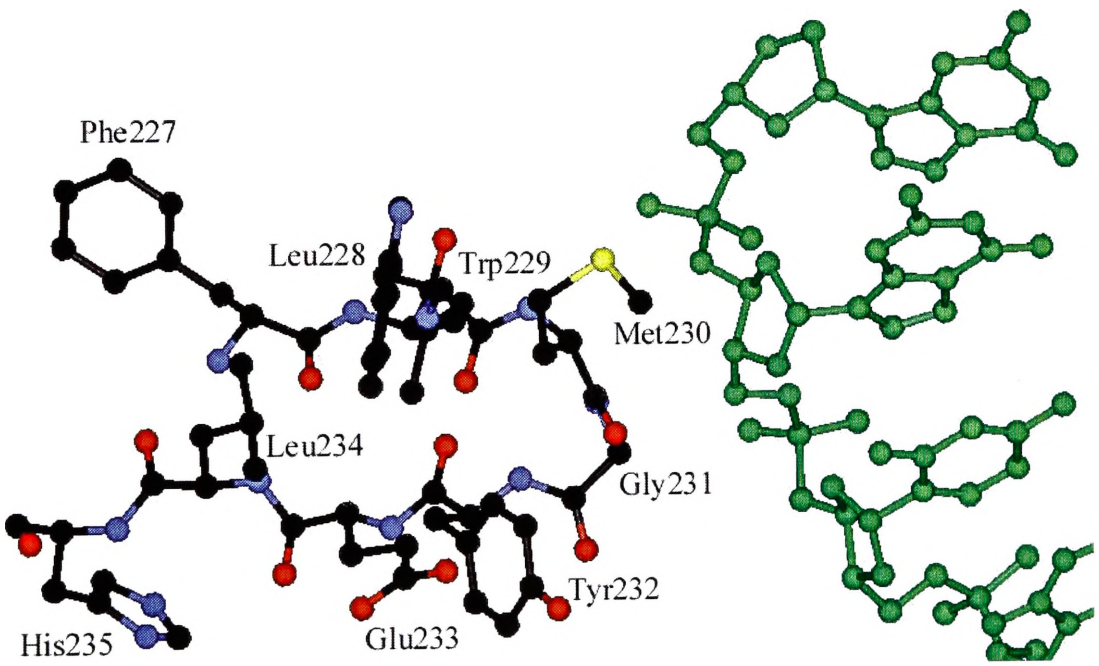


Figure 3.19 Conformation of the primer grip (1rtd) relative to the DNA primer strand (light green). The residues of the primer grip are coloured by atom, oxygen (red), carbon (black), nitrogen (blue), sulfur (yellow).

Interestingly, amino acids Phe227, Trp229, and Leu234 of this region also line the hydrophobic pocket where non-nucleoside inhibitors are known to bind. Consequently it is not surprising that the binding of a NNRTI displaces the primer grip, probably affecting its conformation and/or mobility, and thus its ability to position the 3' terminus of the primer strand at the dNTP-binding site. Alternatively, it may be the mobility of the primer strand itself that is affected by the shifting to the primer grip induced by inhibitor binding. These disturbances in the spatial relationship between the primer terminus, a bound dNTP, and associated divalent metal cation(s) means that an incoming dNTP could still bind without significant alteration of affinity, but the altered relative conformations would not favour the chemical reaction of polymerisation.[139] Furthermore, progressive DNA synthesis requires a translocation movement of about 4-5 Å of the nucleic acid relative to the protein following each nucleotide incorporation.[139] The structural elements of the protein and nucleic acid that participate in this movement are as yet not known, but are certain to include the primer grip. Repositioning of this structural motif would permit it to 'reload' for the conformational changes in the nucleic acid and enzyme that accommodate the next round in the catalytic cycle.[139]

Throughout the entire process of reverse transcription, the primer strand is required to deal with the unique structural demands of several different nucleic acid templates and primers. Almost all of the mutations in the primer grip have a drastic effect on RNA priming activity but not DNA priming activity.[171,173] These effects are not just localised to the polymerase active site, but are as far ranging as to effect the different activities and specificities of the RNase H active site.[172] These observations led to the proposition that interactions of residues in the primer grip with RNA and DNA primers are qualitatively different and determined in part, by RT recognition of the helical structure of the hybrids formed with these primers. A DNA primer annealed to a DNA template, as in the case in (+) strand DNA elongation, will adopt a purely B-form helical structure, while an RNA primer annealed to an RNA template, as in the case in (-) strand initiation, will adopt a strictly A-form structure.[171] RNA-DNA hybrids are neither strictly A- or B-form, having several unusual features, such as theirs helical structure, and a prominent bend in the double helix that distinguishes them from other nucleic acid hybrids. The $tRNA^{Lys3}$ and PPT primers used solely by RT have additional idiosyncrasies which make their recognition and interaction with the enzyme an inordinately uncommon and complex event.[174-176] Minor changes in the primer grip

could be sufficient to disrupt those interactions involved in RNA priming while leaving DNA priming activity relatively unaffected.[171]

3.4.2 Synopsis of the displacement of the primer grip

The primer grip is located at the base of the thumb, on the inner edge of the thumb that is closest to the fingers. It is not actually part of the loops of the protein that form the thumb, but rather lies against the thumb. It is composed of a U-shaped loop, with the arms of the loop beginning at the base of the thumb and extending roughly one-third up the thumb. Residues Met230 and Gly231 make the closest contacts with the sugar-phosphate backbone of the first couple of nucleotides of the primer strand (Figure 3.19), with separation distances of less than 4 Å in the catalytic 1rtd structure. The DDM based superimpositions showed a significant and consistent displacement of the entire primer grip, and not just conformational changes in the sidechains of these residues (Figure 3.20). The residues of the primer grip are listed in Appendix 7.6, and displacement data is tabulated in Appendix 7.8.

The comparison of the unliganded 1dlo structure with the dark green 1hmv unliganded structure, in Figure 3.20, proved to be quite similar, and provided support to compare all other studied structures against 1dlo. The average displacement of the primer grip in 1hmv was 1.13 Å, slightly less than the average displacement of the PAS in 1hmv compared to 1dlo, and on average 2.77 Å less than the displacement seen by the inhibitor-bound structures. The average sidechain displacement was only 1.16 Å, predominantly caused by the gross backbone displacement. The binary RT-DNA structure of 2hmi showed a greater average displacement than the catalytic 1rtd structure, which actually displayed a smaller average backbone displacement of 1.08 Å than the second unliganded 1hmv enzyme. The average sidechain components in 2hmi and 1rtd showed a slight increased difference in displacement from the backbone component, on average 0.31 Å, than seen in 1hmv indicating conformational changes in the primer grip sidechains in addition to the collective displacement of the backbone.

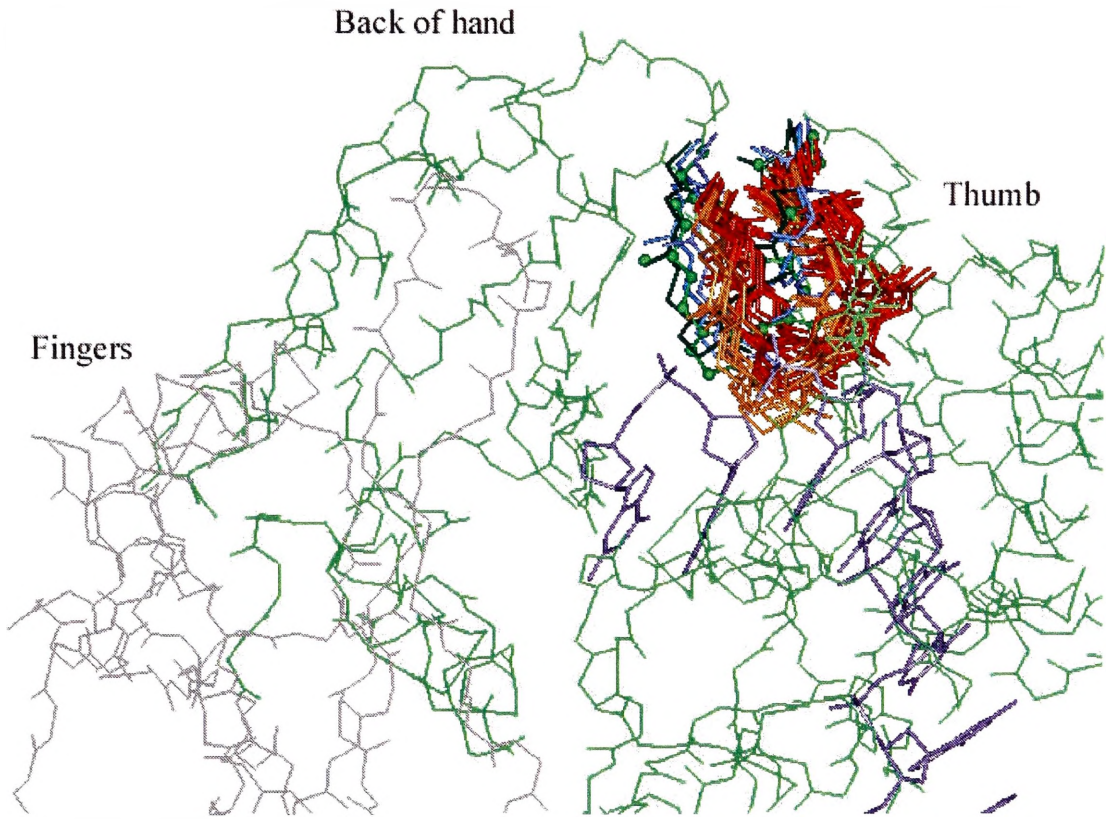


Figure 3.20 An illustration of the displacement of the primer grip upon substrate and inhibitor binding relative to the fingers subdomain, on the left, shown in grey, and the thumb subdomain, on the right. The backbone of 1dlo is shown as a green line trace. Red and orange primer grips represent inhibitor-bound structures, light and dark blue primer grips represent DNA-bound structures. Only the primer strand, in purple, is shown for clarity. This view is the same used in the individual residue analysis, with details of the conformational changes in the text.

The inhibitor-bound structures all showed considerable displacement of the primer grip from the 1dlo structure, and compared to the DNA bound structures. This was most clearly seen in the difference between the multi-coloured, and the black and grey monotone line traces of Figure 3.21, which graphed the scalar displacements of the primer grip from the unliganded 1dlo. The average displacement of the inhibitor-bound structures was 3.90 Å, predominantly composed of the gross backbone displacement, although an average additional displacement of 0.29 Å was seen solely by the sidechain component. Figure 3.21 shows clearly that in the inhibitor-bound structures, the polypeptide region of residues 227-231 was consistently displaced further than the region of residues 232-235. The average displacement for the region 227-231 was 4.82 Å compared to 2.81 Å for residues 232-235. The average displacement values for the inhibitor-bound structures in Table 3.5 also showed differences between the CABM and OCMS research groups, as already seen in the study of the PAS. In the primer grip, the CABM crystal structures were displaced an average 0.67 Å further away from the unliganded 1dlo structure than the crystal structures of the OCMS research group.

average displacement (Å)	1hmv	DNA-bound structures ternary	binary	inhibitor-bound structures total	CABM	OCMS
entire residue	1.13	1.22	1.81	3.90	4.35	3.68
backbone	1.10	1.08	1.59	3.73	4.26	3.49
sidechain	1.16	1.31	1.98	4.02	4.43	3.83

Table 3.5 Summary of the average displacement between the 1dlo unliganded structure and the second unliganded (1hmv), substrate-bound (1rtd-ternary and 2hmi-binary), and inhibitor-bound structures for the entire primer grip, and only the backbone or sidechain component. The three representative average displacements were calculated for all 19 inhibitor bound structures (total), and based on the research groups that derived the crystal structures; six from the CABM research group, and 13 from OCMS research group.

3.4.2.1 Substrate induced conformational changes of the primer grip

The second unliganded structure (1hmv) was similar to 1dlo, in both the backbone and sidechain configurations with no displacement forward toward the primer strand, and only a slight displacement left toward the fingers, most notably for the loop region and β13 strand (residues 230-235). The primer grip in the DNA-bound enzymes (1rtd and 2hmi) was similarly conformed and placed compared to the unliganded enzymes, 1dlo and 1hmv, but not so alike that differences could not be discerned, as seen in the different line traces of Figure 3.22.

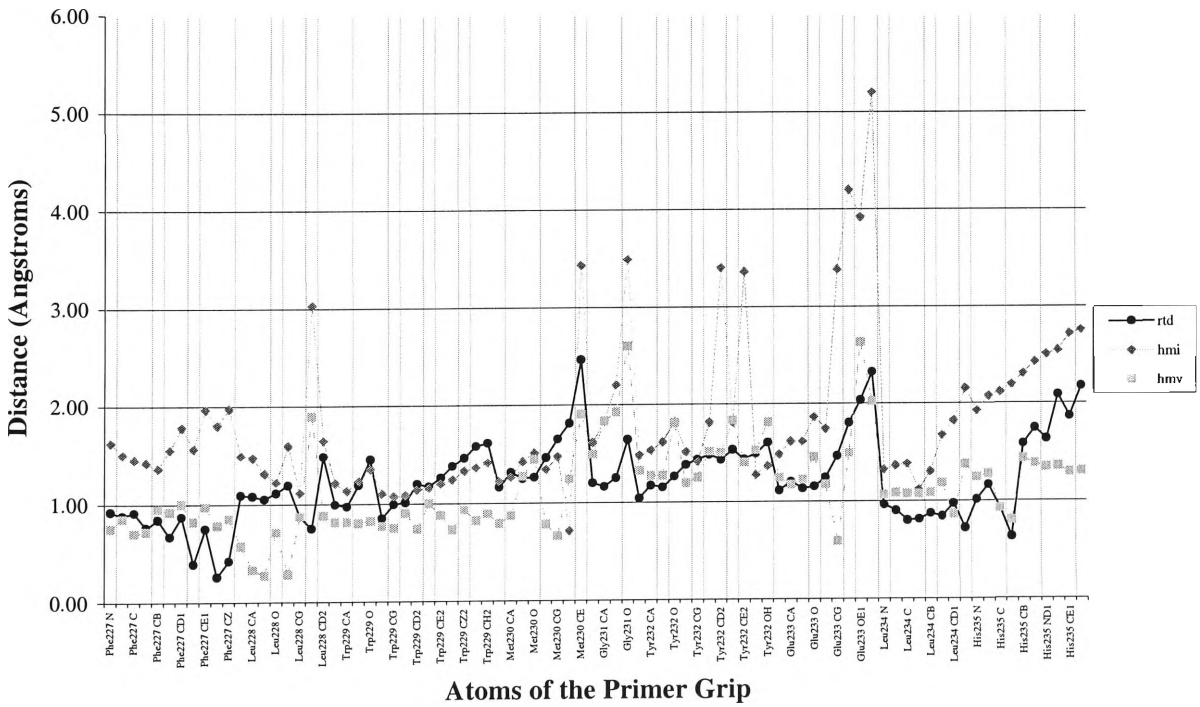


Figure 3.22 Graphical representation of the displacement between the unliganded (1dlo) structure, and the second unliganded (1hmv) and DNA-bound structures (1rtd and 2hmi) for the primer grip.

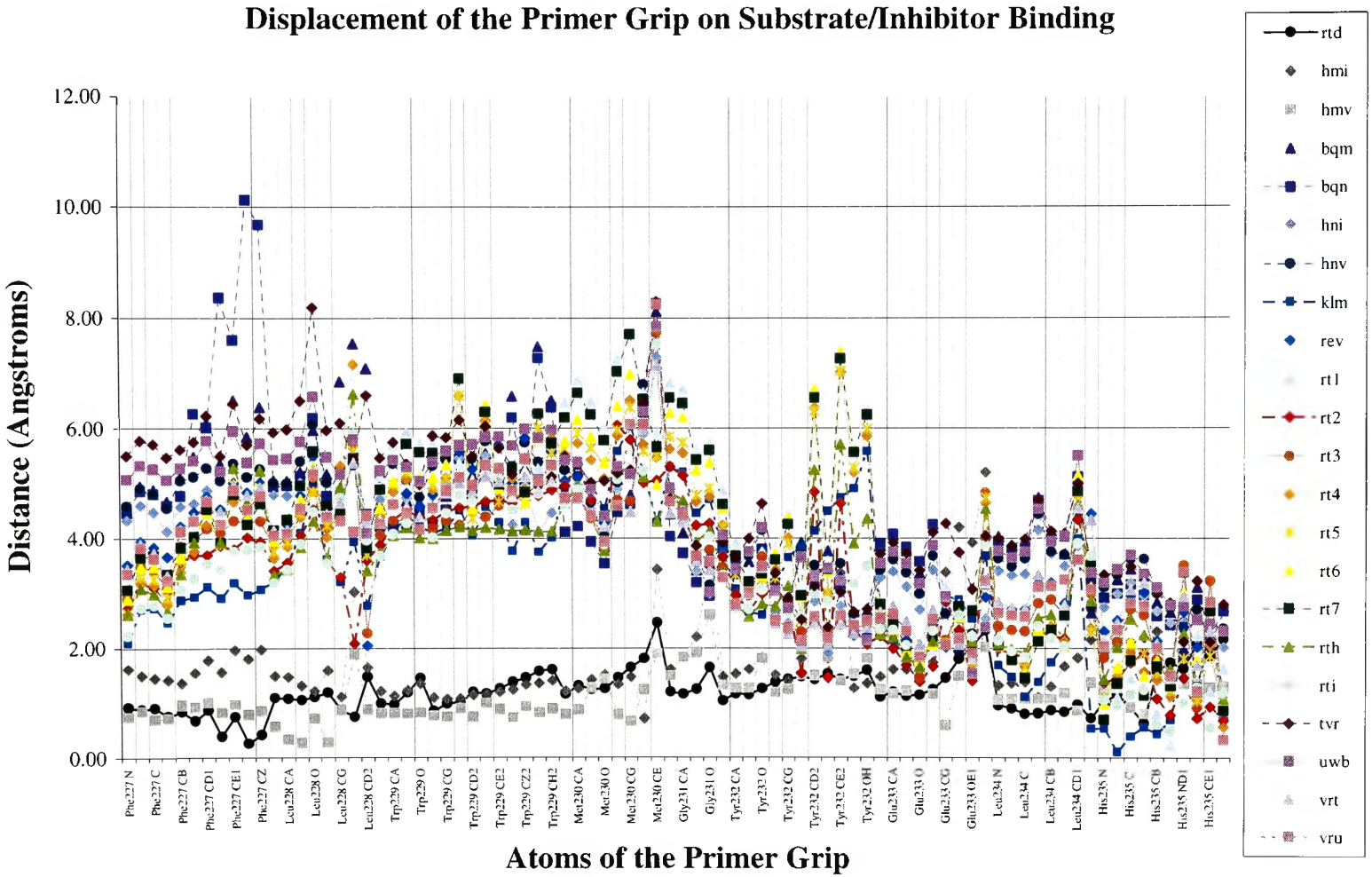


Figure 3.21 Graphical representation of the displacement between the unliganded (1dlo) structure and all RT structures for the primer grip.

DNA binding in the binary structure of 2hmi appeared to cause a small displacement of the primer grip backward away from the fingers or the bulk of the enzyme, as if from a backward retraction of the thumb to accommodate the primer and template strands (see also the extension of the thumb in Figure 3.3). This difference from the unliganded structures, and the catalytic 1rtd structure, is seen in the grey diamond line trace of Figure 3.22, which shows higher than average displacements for Phe227, Gly231, and His235, and the sidechains of Leu228, Met230, Tyr231, and Glu233. It was also seen in the comparison of the dark thin and the dark thick stick structures of Figure 3.23 (left). The catalytic 1rtd structure did not appear to be as affected by DNA binding as the 2hmi structure. The black circular line trace of Figure 3.22 showed only small differences in displacement, not unlike those of the second 1hmv structure. The pale ball-and-stick structure of Figure 3.23 (left) showed virtually no displacement away from the primer strand when comparing it to the dark thick sticks of the 1dlo structure, and only a very small displacement toward the left of the figure, or the back of the hand. No significantly discernible vertical displacement in the plane of the thumb for the primer grip was seen.

3.4.2.2 Inhibitor induced conformational changes of the primer grip

Inhibitor binding induced several concerted displacements of the primer grip from its position in the unliganded state. The conformation of the primer grip loop was relatively unchanged, but the secondary structure of the motif was displaced forward toward the primer strand, even as the rest of the thumb was extended backward away from the primer strand (Figure 3.3 illustrates the movement of the thumb in RT). The two dark structures of Figure 3.23 (right) are distinctly forward of the pale ball-and-stick structure of 1rtd, which closely approximates the structure of 1dlo, seen in the same view and conformation in Figure 3.23 (left). The loop of the primer grip in both inhibitor-bound structures has intruded into the space occupied by the strand in the catalytic DNA-bound structure, most likely having some affect on the positioning of the last two nucleotides of the primer strand.

The primer grip also showed a vertical displacement upward from the palm in the same plane as the thumb, unlike the DNA-bound structures, which showed no significant vertical displacement. The β13 strand of the primer grip in the inhibitor-bound structures showed a slight displacement closer to the fingers subdomain, similar to the

displacement seen in 1hmv, even though the bulk of the primer grip was actually displaced away from the fingers. The β12 strand of the primer grip did not seem to be displaced as such, and so the two strands of the loop appear to tighten. This was not directly manifested in the displacements measured for these residues (Figure 3.21), which appeared to show that residues 227-231 underwent the greatest displacement, though these are only scalar measurements and show no directionality. The direction of the primer grip loop (residues 229-231) also skewered slightly more toward the heel of the hand, or the RNase H subdomain, than the unliganded and DNA-bound structures. This is not so evident in Figure 3.23 (right), but can be seen in Figure 3.24 (right).

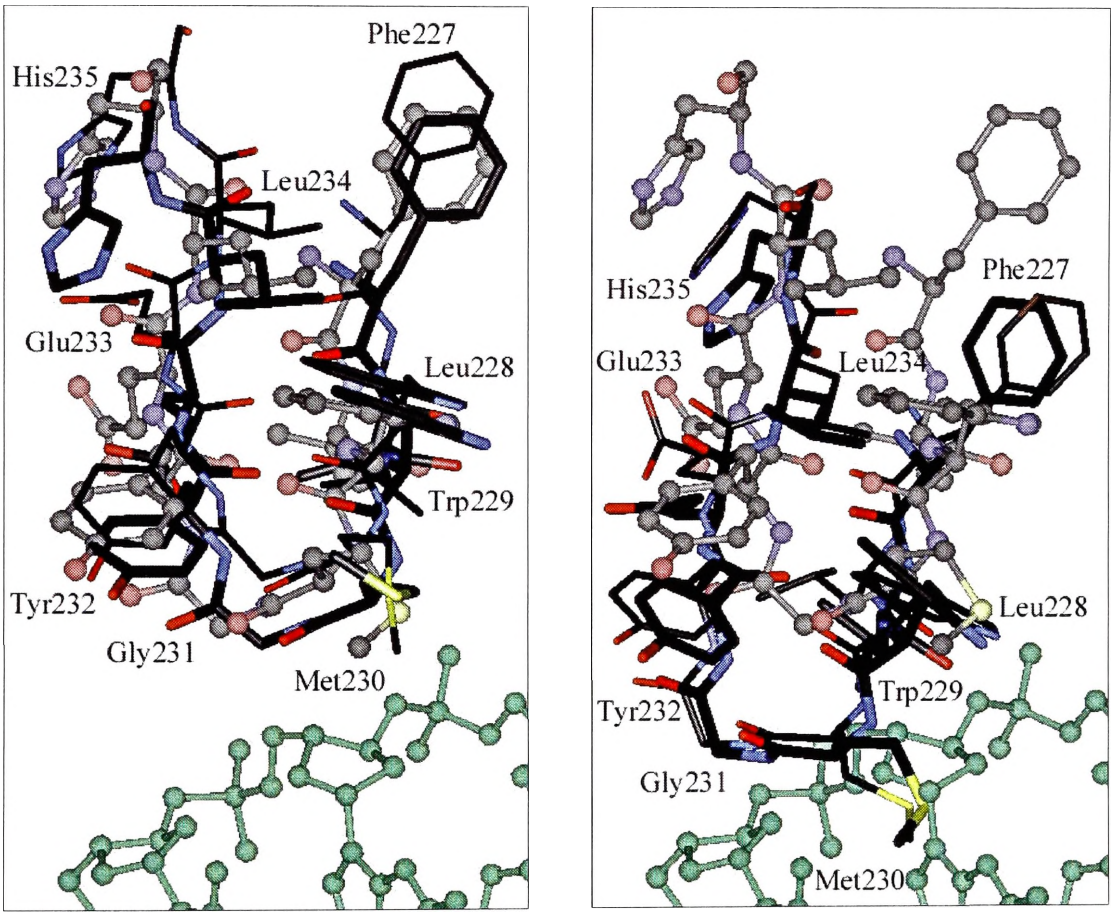


Figure 3.23 (Left) The conformational differences between the unliganded structure 1dlo (thick dark sticks) and the binary DNA structure 2hmi (thin dark sticks), seen relative to the primer grip in 1rtd shown in pale green ball-and-stick (see Figure 3.19 for full legend details). (Right) Inhibitor-bound structures of 1hni (thick dark sticks) from the CABM group and 1vru (thin dark sticks) from the OCMS group, also seen relative to 1rtd

Once again, the differences in the inhibitor-bound crystal structures from each research group could be seen. Those from the CABM group (1bqm, 1bqn, 1hni, 1hnv, 1tvr and 1uwb), coloured in orange in Figure 3.24, showed an additional displacement of the backbone away from the fingers, which was more discernible than that already noted for

the DNA-bound enzymes. The red-coloured inhibitor-bound crystal structures of the OCMS group also exhibited this displacement. Six of the thirteen from this group (1klm, 1rt1, 1rt4, 1rt5, 1rt6 and 1rt7) had a more noticeable displacement away from the fingers. The direction of the primer grip loop also skewered toward the heel of the hand, or the RNase H domain, rather than the fingers region furthest from the thumb, as seen in Figure 3.24 (right). The remaining inhibitor-bound crystal structures of this group also showed a greater displacement away from the fingers than those of the CABM group, but not as extreme as those previously mentioned.

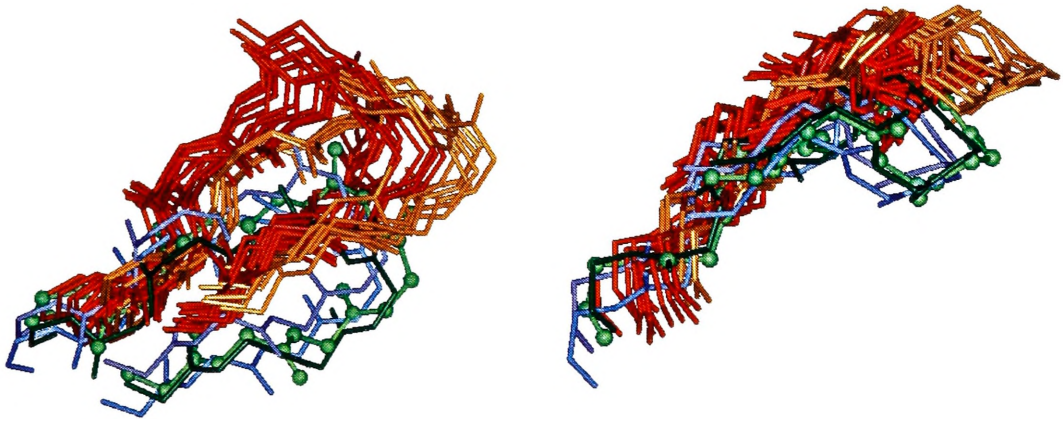


Figure 3.24 Two different conformational views of the displacement of the primer grip upon substrate and inhibitor binding. Both views are from the back of the hand rather than the normal view of Figure 3.1 or 3.3. 1dlo (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red).

Most of the sidechain displacements of the residues of the primer grip were primarily due to the additive effect of the displacement of the backbone. This was reflected in the similar average displacements of the backbone and sidechain (Table 3.5) and the uniformity and smoothness of the displacements seen in the line traces of Figure 3.21. However, some sidechain residues had interesting differences between the unliganded, DNA-bound and inhibitor-bound crystal structures that provided supporting evidence and insight into their function, and were individually investigated.

3.4.3 Conformational changes for phenylalanine 227

This residue's position furthest away from the β12-β13 loop (see Figure 3.19) would have seemed to preclude it from any vital role or interaction with the primer strand. Additionally, mutation of this residue would appear not to have any direct effect on the

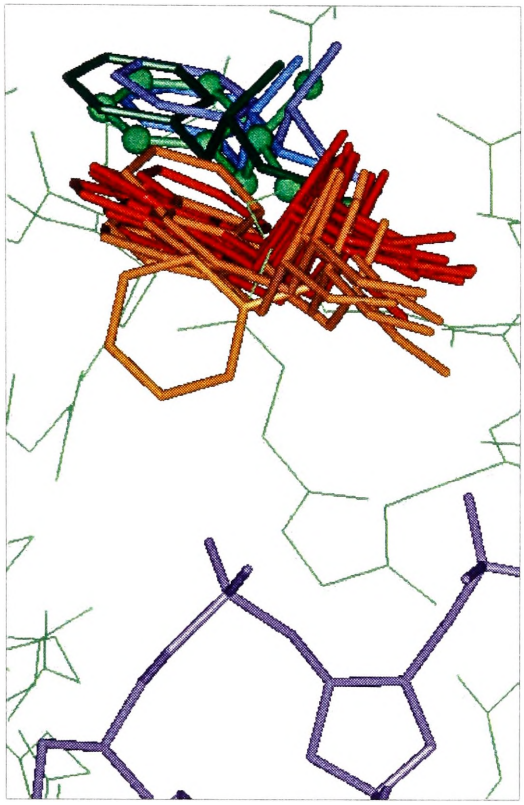


Figure 3.25 The displacement of Phe227 upon substrate and inhibitor binding. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The backbone of 1dlo is shown as a green line trace, the primer strand is in purple. The fingers subdomain is to the left and the thumb subdomain is to the right, as can be more clearly seen in Figure 3.20, which is in the same orientation.

positioning of the 3' terminus end of the primer, yet Phe227 is known to play a role in the specific recognition of the RNA PPT primer, but not the RNA PBS primer.[171]

Aside from the general forward displacement of the whole of the primer grip, no significantly different conformational displacements were seen in the sidechain of Phe227 in either the DNA-bound or inhibitor-bound structures. The grey diamond line trace of Figure 3.22 shows more clearly than Figure 3.25 the greater displacement of the light blue 2hmi structure compared to the dark blue 1rtd structure and the dark green second unliganded structure for Phe227. The catalytic 1rtd structure was much more conformationally similar to 1dlo, as illustrated in Figure 3.23 (left).

The inhibitor-bound structures had a wide spread of displacements in Figure 3.21, from a residue average of 2.82 Å for 1klm to 6.50 Å for 1bqn. The inordinately high average of 1bqn was primarily due to a very different sidechain conformation, seen in Figure 3.25 as the individuated orange phenyl ring, that produced displacements of up to 10.12 Å (Figure 3.21). The individual effect of the displacement of Phe227 in the inhibitor-bound structures was difficult to gauge. The slight differences between the unliganded and DNA-bound structures seen in the sidechain of this residue that pointed to the lower back of the hand, in the direction of the base of the first finger, also revealed little about what role it might play in the catalytically active enzyme. Its distance from actual contact with the primer strand, and its involvement in the structure of the non-nucleoside inhibitor binding pocket, would suggest a supportive role in the architecture of the primer grip, in particular the hairpin

loop that makes contact with the primer strand. Slight changes in its conformation may be sufficient to allow differential accommodation of one primer over another.

3.4.4 Conformational changes for tryptophan 229

A constituent of both the primer grip and the non-nucleoside inhibitor binding pocket, mutation of this residue results in defective DNA synthesis[172] and reduced template-primer affinity.[169] It displayed the most consistent displacement seen of the entire motif, as evidenced by the multi-coloured line traces of Figure 3.21 for the residue. This was also apparent in both DNA-bound structures as well as the second 1hmv unliganded structure in Figure 3.22 where all three line traces stayed below 1.7 Å, and were averaged at 1.1 Å.

The average displacement of the inhibitor-bound structures was 5.0 Å, notably greater than the unliganded and DNA-bound structures, and ranged between an average of 4.1 Å and 5.6 Å. Figures 3.26 and 3.23 (right) illustrate how inhibitor binding induced Trp229 to approach the sugar-phosphate backbone between the third last and the penultimate nucleotide of the primer strand. While Trp229 in the inhibitor-bound structures did not intrude into and occupy the space of the primer strand in the catalytic 1rtd structure, we speculate that it must have some hydrophobic or electronic effect on the positioning of the primer terminus, notwithstanding effects propagated to neighbouring residues. The inhibitor-bound structures showed little difference in the conformation of the aromatic sidechain that pointed toward the middle of the palm. In a few of the structures (1bqm, 1bqn, 1rt4, 1rt5, 1rt6, and 1rt7) the rings of the residue were flipped over by approximately 180° about the C_β-C_γ bond to occupy virtually the

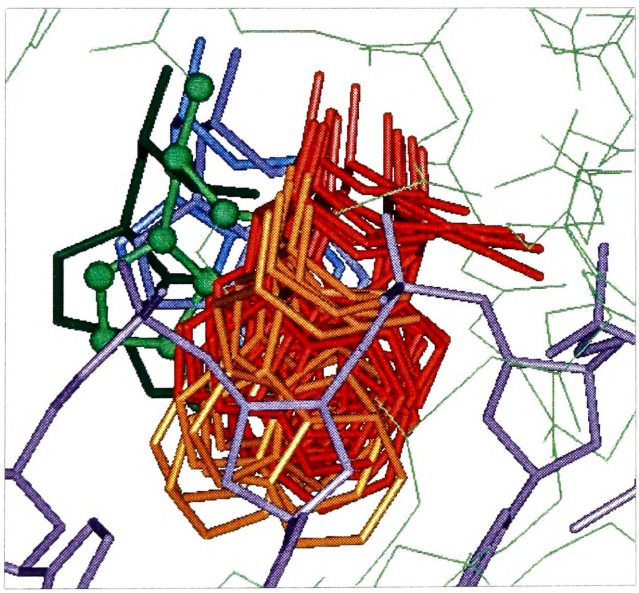


Figure 3.26 The displacement of Trp229 upon substrate and inhibitor binding. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The backbone of 1dlo is shown as a green line trace, the primer strand is in purple. The fingers subdomain is to the left and the thumb subdomain is to the right, as can be more clearly seen in Figure 3.20, which is in the same orientation.

same position [142]. This conformational change would be unlikely to have a steric effect on Trp229 in its role in the non-nucleoside inhibitor binding pocket, but may be quite significant in the bonding interactions it would be capable of with a non-nucleoside inhibitor.

The replacement of Trp229 with other aromatic residues, such as tyrosine or phenylalanine, has been speculated to have the effect of halting the enzyme within a hairpin loop of the PPT-DNA hybrid where the intramolecular base pairing needs to be disrupted in order to synthesise as far as the template 5' terminus.[173] This inhibition affected by only a very slight variation in the structure and nature of this sidechain speaks volumes of the exacting demands of the nature and role played by the primer grip. The π-electron-rich sidechain of Trp229 has been proposed to allow it to participate in stacking interactions with nucleic acid bases and thus interact with and position the primer stand.[169] The greater distance between Trp229 and the sugar-phosphate backbone of the primer strand, compared to Met230 and Gly231, makes it likely that these interactions play only a small role, if any at all. It is more likely that its sidechain provides interactions to stabilise the β12-β13 connecting loop, which constitutes a tight β-turn, in the appropriate architecture of the primer grip.[169] The 180° flipping about the C_β-C_γ bond observed in our results and by other researchers[142] may disrupt the interactions Trp229 is capable of, possibly weakening or totally abolishing its interactions necessary for the stabilisation of the β-turn of the primer grip.

Mutations of this residue and Phe227 have also been observed to have an effect on the fidelity of replication. Some amino acid substitutions reduce enzyme fidelity, while many increase it. The level of fidelity conferred by the natural amino acid sequence is advantageous for the survival of the virus, allowing the enzyme to introduce frequent changes into the genome[177] The inherent flexibility and permissible conformations of these sidechains may directly affect the size and shape of the dNTP-binding pocket, or affect the orientation and/or accommodation of the 3' terminus end in the active site to allow incorrect nucleotides to bind.

3.4.5 Conformational changes for methionine 230

Both the sidechain and backbone of Met230 make contact with the sugar phosphate backbone of the nucleic acid in 1rtd, with distances of less than 4 Å to the second nucleotide of the primer strand.[136] The backbone of the unliganded 1dlo was slightly

closer to the primer strand in 1rtd, while 1hmv and 2hmi are slightly further away. The last atoms of this residue in both DNA-bound structures point toward the primer strand (Figure 3.27) to make close contact with the C_2 of the sugar of the second nucleotide. In 1dlo and 1hmv there is a swivelling about Cγ which resulted in this sidechain pointing into the base of the thumb and the palm respectively, and increasing the separation to the primer strand. This swivelling caused the C_ε atom to be displaced slightly further than the rest of the residue as seen by the increase in the line traces in Figure 3.22 for the last atom of this residue. The differences observed in Met230 between the unliganded and DNA-bound structures were too slight to possibly correlate the conformation to the function or role served in polymerisation or positioning of the primer strand. These differences were also too small to postulate that the residue adopted certain conformations in response to the binding of substrates or to reach a catalytic state.

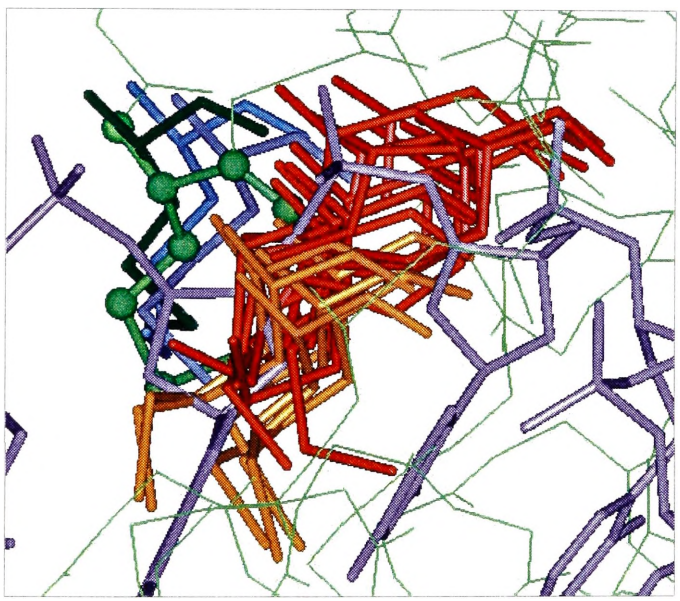


Figure 3.27 The displacement of Met230 upon substrate and inhibitor binding. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The backbone of 1dlo is shown as a green line trace, the primer strand is in purple. The fingers subdomain is to the left and the thumb subdomain is to the right, as can be more clearly seen in Figure 3.20, which is in the same orientation.

Inhibitor binding induced a drastic conformational displacement of this residue forward toward the primer strand, such that Met230 and the neighbouring Gly231 now occupied the space the sugar-phosphate backbone of the second and third nucleotide occupied in 1rtd. The residue was displaced an average of 5.5 Å from 1dlo, with the C_ε atom of inhibitor-bound structures showing displacements of up to 8.2 Å. Figure 3.27 shows clearly the sidechain of Met230 in the inhibitor-bound structures tangling with and occupying the space where the penultimate nucleotide of the primer strand lies in the catalytic 1rtd structure. This displacement indubitably would stall the polymerisation and translocation mechanism of any previously bound primer terminus and incoming dNTP,

or make virtually infeasible the binding of a new primer-template complex for reverse transcription.

3.4.6 Activation and inhibition of the primer grip

The binding of NNRTIs undoubtedly repositioned the β12-β13-β14 sheet, including the primer grip, as several of these residues also comprised the binding pocket for these inhibitors. Despite the much noted hyperextension of the thumb[139] upon inhibitor binding, the displacement of the primer grip revealed by the DDM-based superimposition actually decreased the width of the DNA binding cleft. The hairpin loop is displaced in a direction further down the length of the nucleic acid, away from the dNTP-binding site. Although a distinction can be seen between inhibitor-bound crystal structures from the different research groups (Figure 3.24), this does not lessen the unequivocal and complete displacement of the primer grip in these structures. The large differences seen between the multi-coloured, and the black and grey monotone line traces of Figure 3.21, especially for residues 227-231, illustrated clearly the magnitude of the displacement of the primer grip upon inhibitor-binding.

The gross displacement of the thumb (Figure 3.3) is speculated to be necessary for the accommodation of nucleic acid hybrids in the DNA-binding cleft. Curiously, despite its close proximity at the base of thumb, the primer grip in both DNA-bound structures showed virtually none of this extension to a more 'open' configuration. The slight differences in displacement of the primer grip between the unliganded and DNA-bound structures do little to shed light on structural changes that may be necessary to achieve a catalytic conformation. It may be that the minor shifting of the primer grip, in 1rtd and 2hmi compared to 1dlo, was a case of local structural changes to conform to the structure of the bound DNA primer. The binding of different RNA primers necessary for the different initiation stages of reverse transcription would presumably require a greater flexibility of the primer grip, as these primers do not have the basically linear helical structure seen in dsDNA. This flexibility may be manifested as more significant structural and conformational changes of the primer, which may provide a rationale for experimental results showing differing polymerisation and dissociation rates between RNA and DNA primers. The slower extension rate of RNA primers may be due to the unique secondary and tertiary structure of the $tRNA^{Lys3}$ and PPT primer complexes,[178] and the conformational changes in the primer grip necessary when accommodating and

retaining these nucleic acids in the binding cleft. In the absence of different RNA-primer-bound crystal structures, the magnitude of this flexibility, and the conformational changes expected, can only be speculated at this time.

The DDM-based superimposition has demonstrated that inhibitor binding induces significant displacement of the primer grip, in the opposite direction to the hyperextension of the thumb. The primer grip intruded into the nucleic acid binding cleft, but this did not result in a decrease in its separation from the residues of the PAS. The distance between the C_α of Gly231, in the loop of the primer grip, and the C_α of the residues of the catalytic triad was measured and analysed in Figure 3.28 (data tabulated in Appendix 7.9).

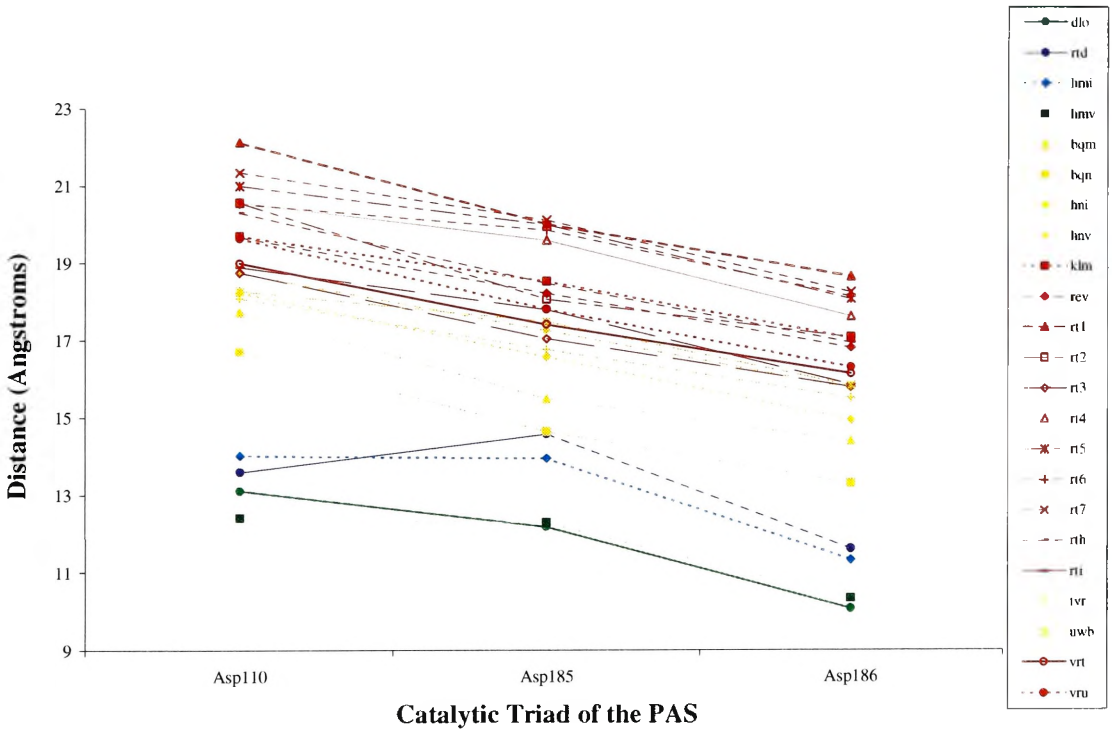


Figure 3.28 Graphical representation of the distance between C_α Gly231 of the primer grip and C_α of the residues of the catalytic triad in all of the crystal structures. 1dlo (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red).

The smallest distances were measured in the unliganded structures, followed by the DNA-bound structures; the difference between the unliganded and substrate-bound structures was only an average 1.45 Å. The inhibitor-bound crystal structures also displayed conformational changes in the primer grip, which was forced forward into the nucleic acid binding cleft, but in a direction which actually increased its distance from

the active site residues. The orange coloured lines traces of the CABM research group crystal structures in Figure 3.28 averaged a 16.44 Å separation between the primer grip and the PAS, while the red coloured OCMS structures had an average separation of 18.68 Å. This was a reversal of the average overall displacements of the primer grip in the inhibitor-bound structures seen in Table 3.5, where the crystal structures of the CABM research group displayed a greater average displacement from the unliganded 1dlo structure.

The significance of the data in Figure 3.28 cannot be extrapolated beyond the observation that inhibitor binding changes the spatial relationship between the primer grip and the catalytic aspartates, increasing it by an average of 4.79 Å from the substrate-bound structures. This is certain to effect the communication necessary between the primer grip and the aspartates, as it positions the 3'OH primer terminus in relation to the incoming nucleotide, the Mg^{2+} ions, and the catalytic residues of the PAS. While the horizontal displacement of the primer grip would seem to impede the binding of any nucleic acid, by occupying part of the binding cleft, the vertical displacement may compensate for this loss of space, and rationalise the observation that nucleic acids can still bind in the presence of NNRTIs. The conformation of the primer grip in an inhibitor-bound structure, or the ability of the template-primer complex to bind to the enzyme may depend on which event occurred first, inhibitor or template-primer binding.

The binding of NNRTIs has been demonstrated to have a profound affect on both the primer grip motif and the thumb subdomain of RT (Figure 3.3 and 3.24). The binding of DNA substrates had a smaller effect on the primer grip, but an analogous effect on the thumb subdomain as inhibitor binding. These induced structural changes were in opposite directions, and those of the thumb subdomain warranted some further study by reason of its structural closeness to the previously studied primer grip. In the DNA-bound structures, the thumb subdomain underwent a hyper-extension to accommodate the binding of nucleic acids in the major cleft of the enzyme is seen in the blue C_α trace of 1rtd in Figure 3.30 compared to the green trace of the unliganded 1dlo. The extension from the 'closed' position in the unliganded structure was investigated by measuring the displacement from the 1dlo unliganded structure to all other RT structures at three positions in the thumb subdomain (illustrated in Figures 3.29 and 3.30, data tabulated in Appendix 7.10). DNA-bound structures showed an extension of

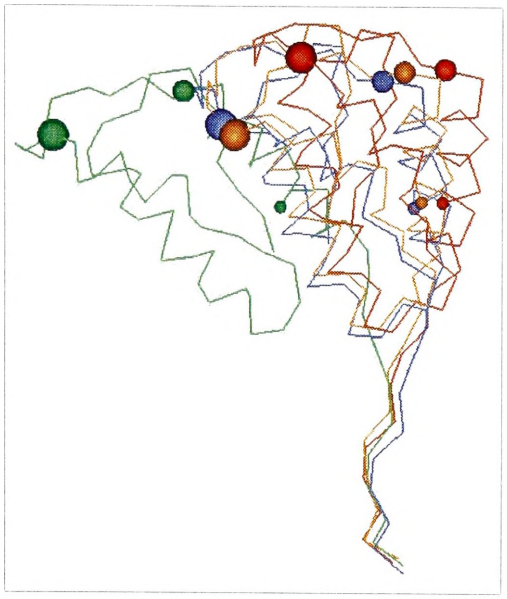


Figure 3.29 Cα trace of the thumb subdomain of RT with the Cα atoms of residues 286, 296, and 306 in all structures represented by large, medium and small balls, respectively, representing various parts of the thumb. 1dlo (green), 1rtd (dark blue), 1hni (orange), and 1vru (red).

about 16.8 Å in 1rtd and 18.2 Å in 2hmi. The inhibitor-bound structures demonstrated a greater extension of the thumb than previously seen in the DNA bound thumb subdomains, and was most notable in the red coloured structures from the OCMS research group (Figure 3.30).

In Figure 3.29 the differences in the C_α line traces are only slight, with the orange-coloured CABM structure (1hni) appearing almost the same as the dark blue DNA-bound structure (1rtd). However the consistent trends of Figure 3.30 show clearly the greater extension of the thumb in inhibitor-bound structures, on average 20.32 Å.

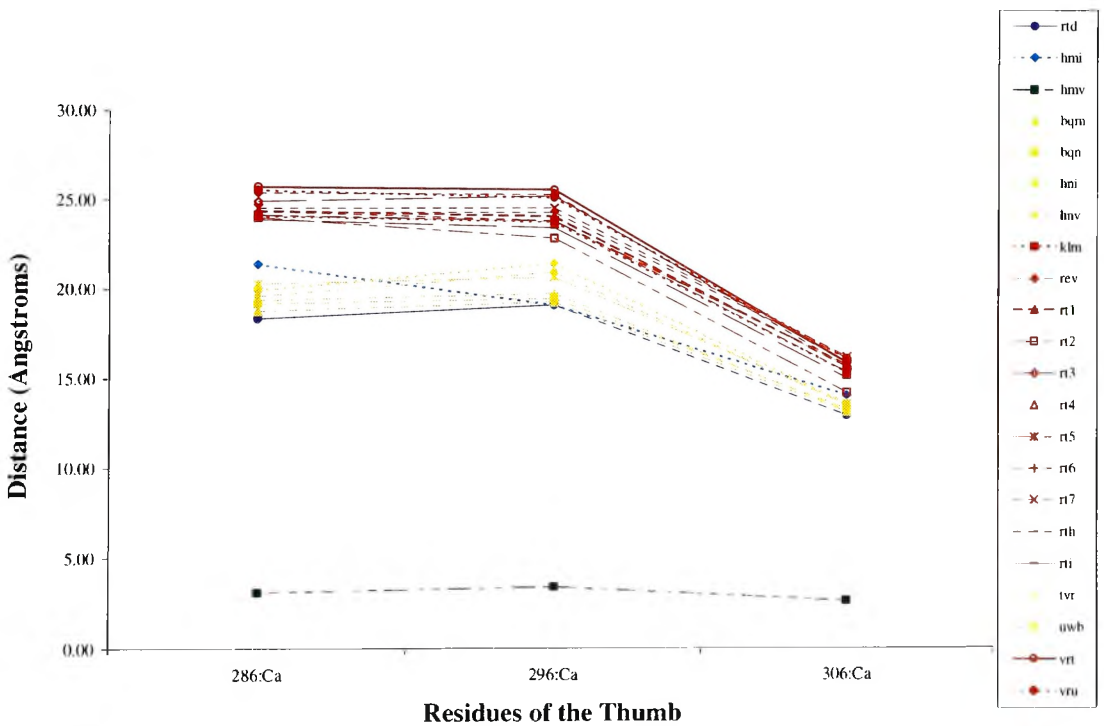


Figure 3.30 Graphical representation of the distance between the unliganded (1dlo) structure and all other RT structures for three residues representing various parts of the thumb (Figure 3.29). 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red).

The OCMS structures averaged a thumb displacement of 21.51 Å, while the CABM group displacement was only 17.75 Å, just marginally greater than the average displacement of the DNA-bound thumb domains.

The greater extension of the thumb in inhibitor-bound structures as compared to substrate bound structures may affect inhibition of polymerisation. This may not only be in terms of affecting its mobility as suggested by the 'molecular arthritis model' of non-nucleoside inhibition (Section 3.13), but also possibly by influencing the stability of the primer grip which lies in close contact with the thumb. The movement of the primer grip forward into the DNA binding cleft is not followed by a forward movement of the thumb, but rather a backward hyperextension that may abolish or weaken stabilising interactions between the primer grip and the thumb subdomain. Further detailed analysis of the structural relationship between the primer grip and the thumb subdomain in future may provide further insight into this speculation.

Unlike the polymerase and RNase H active site, which catalyse chemical reactions, the primer grip serves more as a structural device to position and translocate the primer strand and incoming dNTP for each round of polymerisation. This is not to deprecate its function in reverse transcription, but rather clearly delineate the nature of its role. Subtle differences in the primer grip architecture, brought about by substitution of certain residues or induced by inhibitor binding, have profound detrimental effects on all aspects of all of the enzymatic activities of RT. The enzyme's ability to discriminate between oligoribo- and oligodeoxyribonucleic acid primers in general, and $tRNA^{Lys3}$ and PPT primers in particular, is affected. This may stem from its altered position deeper in the nucleic acid binding cleft, or may be a result of reduced flexibility restricting its ability to perform dynamic movements such as translocation of the primer strand and adaptation to different primers and nucleotides.

The significance of the primer grip for enzyme activity and virus infectivity is reason enough for its intense study; the participation of several of its key residues in the formation of the non-nucleoside inhibitor binding pocket augments this need further. Residues such as Trp229, make ideal targets for drug design, as they form part of the binding pocket of non-nucleoside inhibitors, are vital for the integrity of the primer grip, and their mutation does not result in a high resistance profile to NNRTIs.[179]

3.5 The template grip

3.5.1 The template grip and its role in reverse transcription

The 'template grip', consisting of the β4, αB, β5a, and the β8-αE connecting loop portions of the p66 palm and finger, is closely associated with the template strand as it is first reverse transcribed and then hydrolytically degraded by HIV-1 RT.[168] Not as much data has been generated, collated or interpreted about the functions of the residues of the template grip, as for the PAS and primer grip. This may be due to an assumption that the template grip plays a less active role in reverse transcription by reason of its distance to the PAS and the 3'OH of the primer terminus where incoming nucleotides add to the growing DNA strand. Additionally its distance, and possible irrelevance to the non-nucleoside inhibitor binding pocket has meant that little effort has been exerted to understand the conformational changes that occur in the template grip upon inhibitor binding, and how these changes may promote inhibition of the enzyme.

What mutational data of the residues of the template grip is available is often fragmentary, with seemingly random residues investigated and their affect on polymerase or RNase H functions noted without any attempts at a complete or even general mechanism of action for the template grip.[180-182] What is known is that many of the mutations that confer resistance to nucleoside inhibitors are clustered in the fingers and palm subdomains.[180,181] Presumably, interactions between the β3-β4 region and the template affect the ability of the enzyme to discriminate nucleoside analogue drugs from natural dNTP substrates without compromising catalytic activities essential for viral replication.[33]

Regions of the fingers subdomain have been implicated in polymerisation processivity,[160,180,181,183] which refers to the number of nucleotides added to a nascent DNA strand before the enzyme dissociates from the template primer complex.[163] HIV-1 RT is known to display a moderate-to-poor degree of polymerase processivity, which may have been evolutionarily conserved to facilitate the strand transfer reactions (Section 3.1.2). Several factors may contribute to low levels of processivity; weak template-enzyme interactions, secondary structure, and other sequence-specific structural variations along the nucleic template.[183] A greater understanding of the structure of the template grip, and how substrate and inhibitor binding affect its

constituent residues should provide more insight into its function and the exact role it plays in processivity.

The template grip is composed of three segments of polypeptide as seen in Figure 3.31. The first section runs horizontally across the middle of the fingers formed by the β4 and αB secondary elements. The second section composed of the β8-αE connecting loop lies across the middle of the palm, while the third section, formed by β5a, runs across the bottom of the fingers where the fingers meet the palm.

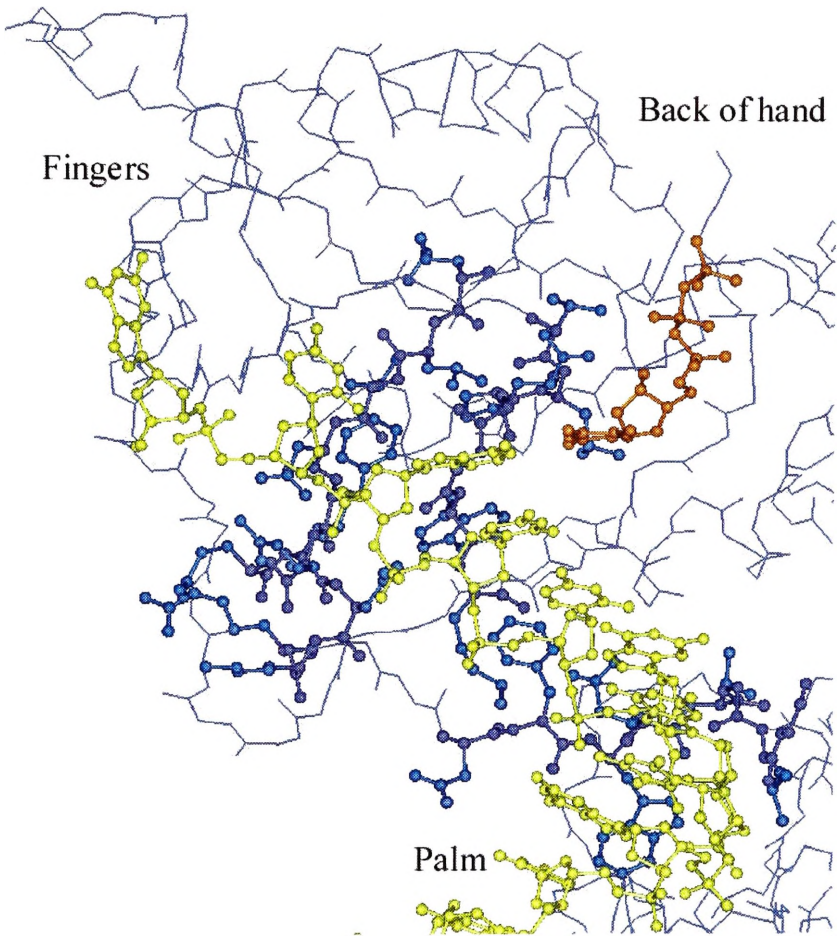


Figure 3.31 Conformation of the three segments of the template grip (1rtd) relative to the DNA template strand (yellow) and the incoming nucleotide (orange), viewed from above the hand, the primer strand has been removed for clarity. Two segments run across the middle and the bottom of the fingers, and third segment runs across the middle of the palm. The backbone of 1rtd is in dark blue lines, the backbone of the template grip is in dark blue ball-and-stick, and the sidechains are in light blue ball-and-stick.

In the catalytic 1rtd structure, several residues of the template grip were observed to make possible contact with the template strand of the DNA, as well as the incoming orange dNTP as seen in Figure 3.31. The sidechain of Gln151 comes within 3.0 Å of the incoming nucleotide that has been positioned to pair with the templating base. The

base of the nucleotide awaiting based-pairing with the incoming dNTP, *n*, is approximately 3.5 Å away from the sidechain of Leu74. The sidechains of Asp76 and Asp78 lie 2.7 Å and 3.8 Å respectively away from the sugar phosphate backbone between *n* and *n*-1, the template nucleotide already base-paired to the last nucleotide of the primer strand. Lys154 makes contacts with the sugar-phosphate backbone between *n*-1 and *n*-2, while Glu89 lies 3.7 Å from the sugar of the nucleotide *n*-2.

3.5.2 Synopsis of the displacement of the template grip

The template grip is composed of three segments of polypeptide. The first section runs across the middle of the fingers (residues 73-83), the second section lies across the middle of the palm (residues 86-93), and the third section runs across the bottom of the fingers (residues 148-154) where the fingers meet the palm. The residues of the template grip are listed in Appendix 7.6, and displacement data is tabulated in Appendix 7.11.

The comparison of the 1hmv unliganded structure revealed little difference with the 1dlo structure in the backbone of the structures, with an average displacement of only 0.68 Å. The sidechain component showed more of a displacement than the backbone, shifting an average of 1.16 Å from the conformation of 1dlo. The greatest differences in the sidechain conformations of the two unliganded structures were noted at residues Val90, Gln91, and Gln151, as can be seen in the light grey square line trace of Figure 3.33 and detailed in Section 3.5.4 and 3.5.5. With the exception of these few residues, both the backbone and sidechains of the two unliganded structures were virtually indistinguishable, and served to verify and substantiate later comparisons with the 1dlo unliganded structure.

The results of the DDM-based superimposition showed that there were several small differences in the conformations of the two DNA-bound structures compared to the unliganded structures, despite the almost identical average displacement values for 1rtd and 2hmi, differing by an average of only 0.07 Å. The values in Table 3.6 for the ternary and binary structures were slightly misleading as they encompassed the conformational changes of the three separate segments of the template grip, which each behaved quite differently in terms of substrate-induced conformational changes. The average displacement of the 1rtd catalytic ternary structure was observed to be marginally greater than that of the 2hmi structure in both the backbone and sidechain

components. This was in contrast to the average displacements for these structures seen in the PAS and the primer grip, where the binary structure had displayed notably greater average displacements than 1rtd.

The inhibitor-bound structures displayed an average displacement of the template grip of 4.13 Å, from the conformation in the 1dlo unliganded structure. Figure 3.32 shows the difference of approximately 2.6 Å between the multi-coloured, and the black and grey monotone line traces, representing the inhibitor-bound and substrate-bound or unliganded structures respectively. Consistent displacements of varying magnitudes were seen throughout most of the three segments of the template grip, separated by spaces in Figure 3.32, with larger displacements seen in the sidechains of the residues compared to the backbone. The least consistent residues of the template grip in the inhibitor-bound structures were residues 89-93 of the palm segment that showed displacements as great as 14.38 Å in a leucine sidechain of 1rt3 and as little as 2.24 Å in 1rt2. The continuing trend of greater average displacements in inhibitor-bound crystal structures of the CABM research group compared to the OCMS group was observed in Table 3.6, with CABM crystal structures being displaced 0.85 Å further from the unliganded 1dlo structure than those from the OCMS group.

average displacement (Å)		DNA-bound structures		inhibitor-bound structures		
	1hmv	ternary	binary	total	CABM	OCMS
entire residue	0.94	1.83	1.76	4.13	4.71	3.86
backbone	0.68	1.69	1.61	3.83	4.33	3.60
sidechain	1.16	1.95	1.90	4.38	5.02	4.08

Table 3.6 Summary of the average displacement between the 1dlo unliganded structure and the alternate unliganded (1hmv), substrate-bound (1rtd-ternary and 2hmi-binary), and inhibitor-bound structures for the entire template grip, and only the backbone or sidechain component. The three representative average displacements were calculated for all 19 inhibitor bound structures (total), and based on the research groups that derived the crystal structures; six from the CABM research group, and 13 from OCMS research group.

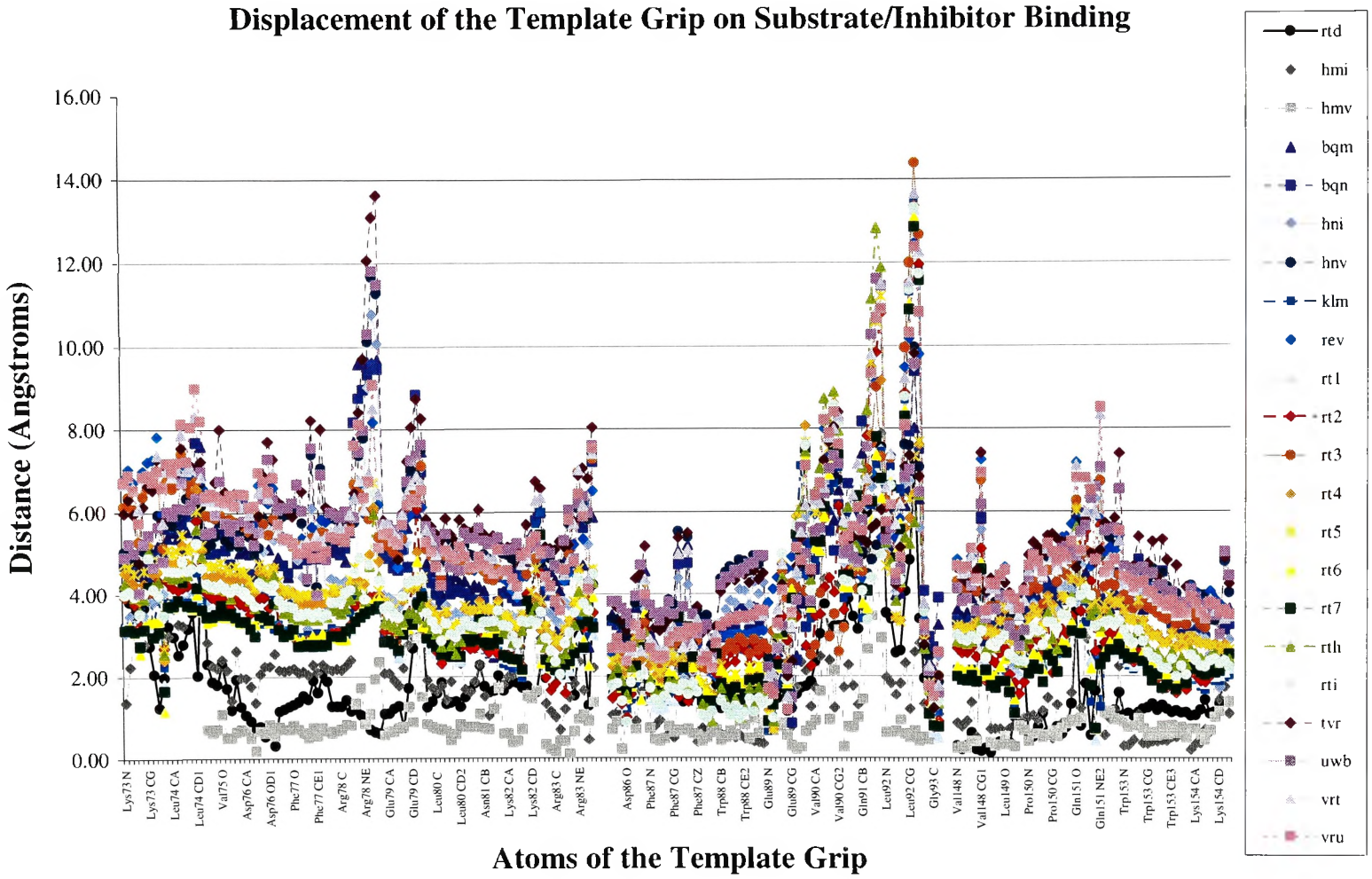


Figure 3.32 Graphical representation of the displacement between the unliganded (1dlo) structure and all RT structures for the template grip.

3.5.2.1 Substrate induced conformational changes of the template grip

In Figure 3.33, the DNA-bound structures of 1rtd and 2hmi did not appear to show great similarity to the unliganded structures in the template grip, but closer observation of Figure 3.34 (left, top and bottom) revealed some residues that were quite comparable in conformation and positioning. Specifically, in the segment of the template grip running across the bottom of the fingers (residues 148-154) there was virtually no difference in the template grip between the DNA-bound enzymes and the unliganded enzyme (Figure 3.34 left, bottom). The black circle and grey diamond line traces of Figure 3.33 in this region showed fairly consistent average displacements of 1.1 Å, with only a few deviations in the occasional sidechain, such as Gln151 in 2hmi, displaced 4.85 Å.

Most of the segment of the polypeptide across the middle of the fingers, furthest from the thumb, showed uniformity between 1rtd, 2hmi and 1dlo. This portion of the segment was composed of residues 76-83. The remainder of this segment, 73-75, closest to the thumb, had the DNA-bound structures displaced slightly lower in the plane of the fingers, with average displacements of about 2.8 Å seen in Figure 3.33 for these residues in the black circle and grey diamond line traces. The thin dark stick structure of 2hmi in Figure 3.34 (left, top) was, overall, slightly displaced back into the fingers, as if from an opening or extension of the hand, while pale ball-and-stick structure of 1rtd was slightly displaced forward of the thick dark stick structure of the unliganded enzyme. This was also seen in the dark and light blue backbone traces of Figure 3.35 (left). In 1rtd this could be speculated to be a movement associated with clamping of the template-primer grip to interact with the nucleic acid as it is reverse transcribed.

Along the palm of the hand, the DNA-bound structures were very similar to 1dlo, as illustrated by the almost overlapping structures for these residues in Figure 3.34 (left, top), and the smaller valued black and grey line traces of Figure 3.33. The exceptions to this were the residues Val90 and Gln91, where the polypeptide curved slightly toward the PAS, before the last few residues curved back to a similar position as the unliganded enzyme (Figure 3.35). This deviation in conformation, was not totally evident in Figure 3.34 (left, top), but was manifested in the greatest displacement of about 7.5 Å from the unliganded structure in the entire template grip in Figure 3.34. No discernible vertical displacement in the plane of the palm was noted for either DNA-bound structure.

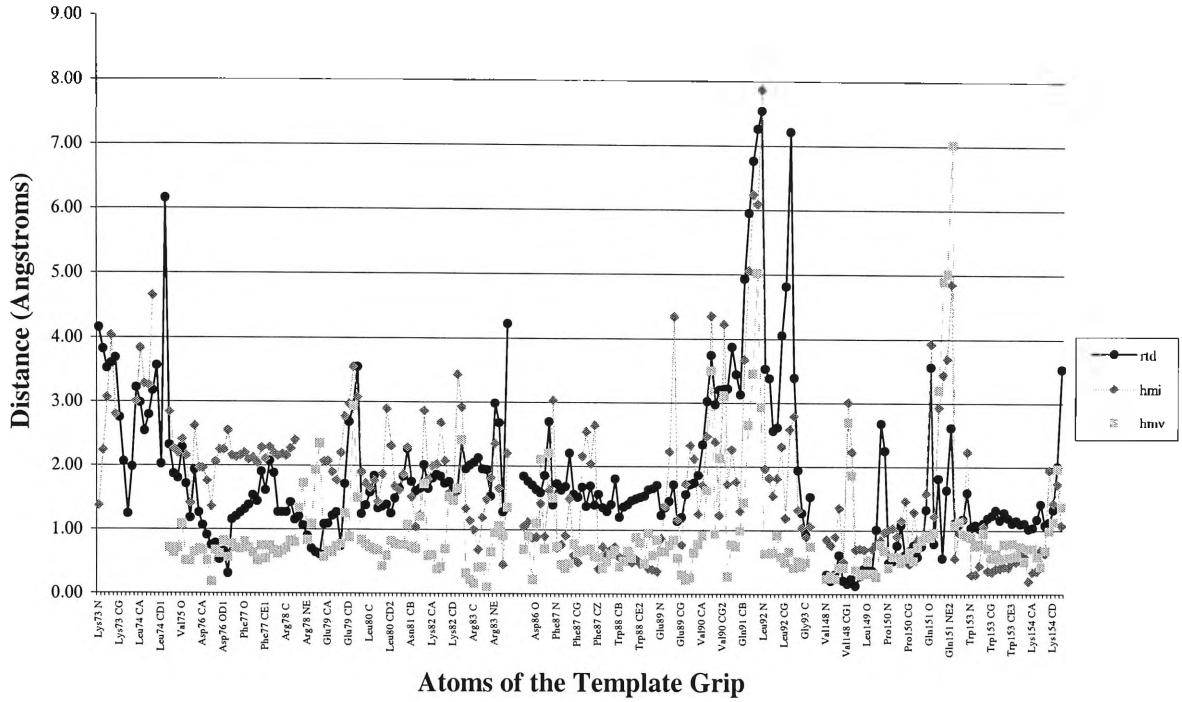


Figure 3.33 Graphical representation of the displacement between the unliganded (1dlo) structure, and the second unliganded (1hmv) and DNA-bound structures (1rtd and 2hmi) for the template grip.

3.5.2.2 Inhibitor induced conformational changes of the template grip

The binding of non-nucleoside inhibitors induced unambiguous conformational changes in the three separate segments of the template grip. The clustering of multi-coloured line traces for the inhibitor-bound structures was distinct from the lower placed black and grey line traces of the unliganded and DNA-bound structures in Figure 3.33. Consistent displacements were seen across the sequences of the template grip, with the exception of residue 78 and residues 89-93, which showed much greater displacement and lacked the clustering effect seen in residue regions such as 80-82 and 86-88.

Varying displacements back into the fingers were seen for the section of the primer grip that runs across the middle of the fingers (Figure 3.34, right, top, Figure 3.35). This displacement varied, with a number of crystal structures (1rt1, 1rt2, 1rt6, 1rt7, 1rth and 1rti) clustering close to the light blue 2hmi DNA-bound structure in Figure 3.35 (left), and the remainder being displaced further back into the fingers. A smaller displacement downward, in the plane of the fingers was also seen in this segment of the template grip.

In the segment of the template grip that runs across the palm, residues Asp86 to Trp88, near the fingers, displayed a slight displacement down into the plane of the palm (Figure 3.35, above). The residues Val90 and Gln91 of all the inhibitor-bound enzymes, with the exception of 1rt2 and 1rt3, showed a greater curvature toward the polymerase active site than the slight curving displacement already previously noted in the DNA-bound enzymes. The dark black and grey structures of Figure 3.34 (top, right) and the orange and red structures of Figure 3.35 (left) show clearly how the last couple of residues, Leu92 and Gly93, of this segment of the template grip curve back to a similar position to the unliganded and DNA-bound structures. This segment of the template grip was also displaced vertically upward from the plane of the hand; on average 2-3 Å, possibly affecting the template strand that runs through the nucleic acid binding cleft.

The final segment along the bottom of the fingers in the inhibitor-bound structures behaved similarly to the section of the template grip along the middle of the fingers. There was a general displacement back into the fingers, away from where the template strand of DNA would contact the fingers subdomain (Figure 3.34, left), and a small downward displacement in the plane of the fingers (Figure 3.34, right, bottom). An additional slight displacement was noted where the polypeptide of the inhibitor-bound structures was displaced toward the thumb, in the plane of the fingers, that is, toward the first finger and away from the smallest finger. This can be seen in Figure 3.34 (right, bottom) where the thick and thin dark structures are shifted upward in the illustration.

The larger number of residues of the template grip and the smaller body of literature on this motif make it difficult to confidently correlate the conformational changes observed by the DDM-based superimposition to functional roles of the residues, or how these roles may be stalled by inhibitor binding. Several residues of particular interest were analysed in more detail to provide some basis for speculation to the roles as residues of the template grip.

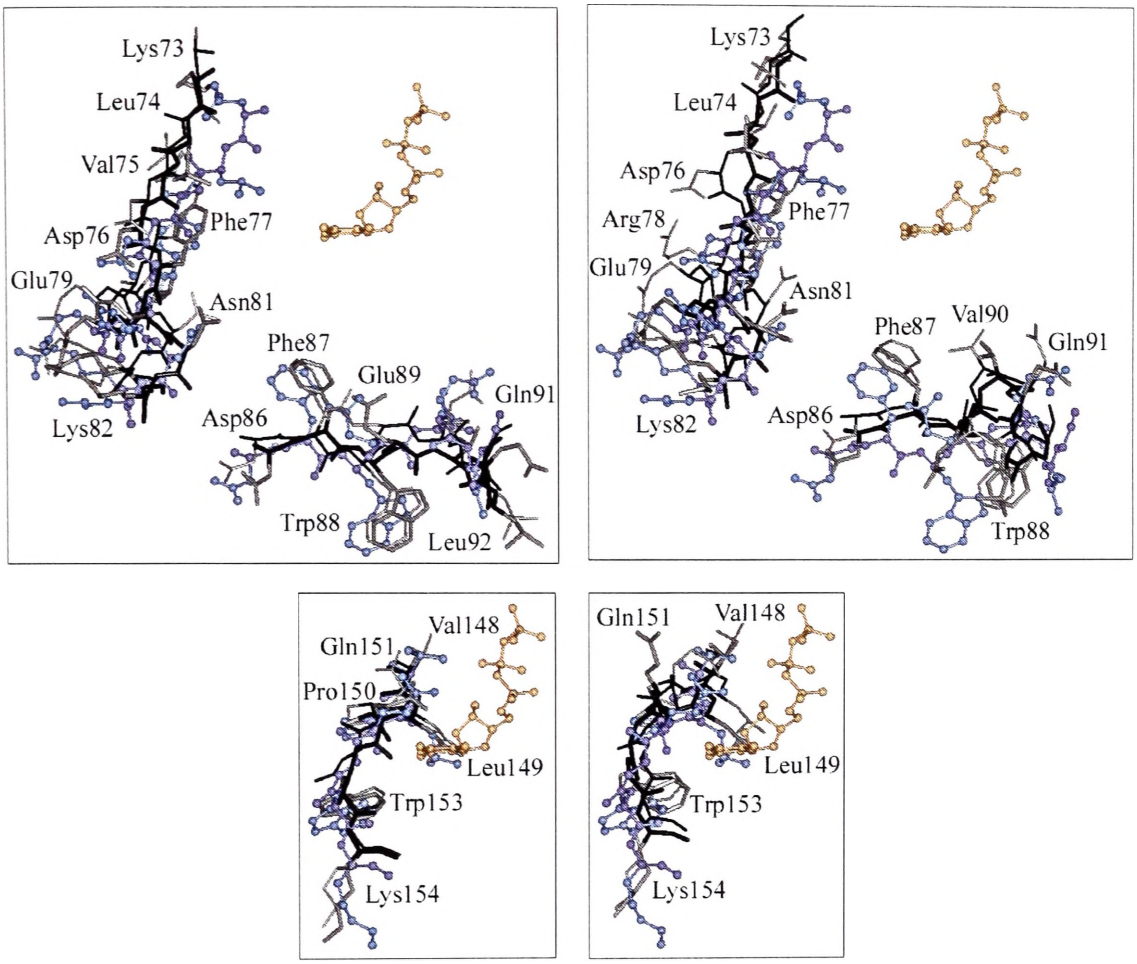
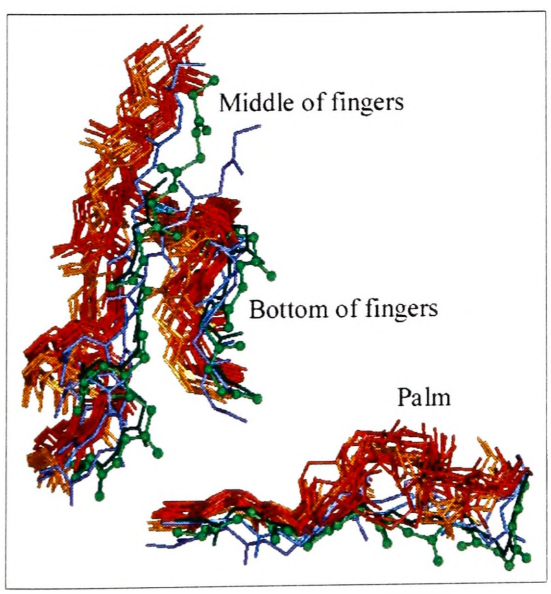


Figure 3.34 (Left, top and bottom) The conformational differences between the unliganded structure 1dlo (thick dark sticks) and the binary DNA structure 2hmi (thin dark sticks), seen relative to the template grip in 1rtd shown in pale ball-and-stick. (Right, top and bottom) Inhibitor-bound structures of 1hni (thick dark sticks) from the CABM group and 1vru (thin dark sticks) from the OCMS group, also seen relative to 1rtd. (Top) The middle of the fingers and palm segments of the template grip. (Bottom) The bottom of the fingers segment of the template grip. The conformation of all figures is the same as in Figure 3.31; the incoming nucleotide is in orange, the backbone of 1rtd is in pale dark blue ball-and-stick, and the sidechains are in pale light blue ball-and-stick; the backbone of 1dlo, 2hmi, 1hni, and 1vru are in dark black, and the sidechains are in dark grey. This view is the same used Figure 3.31



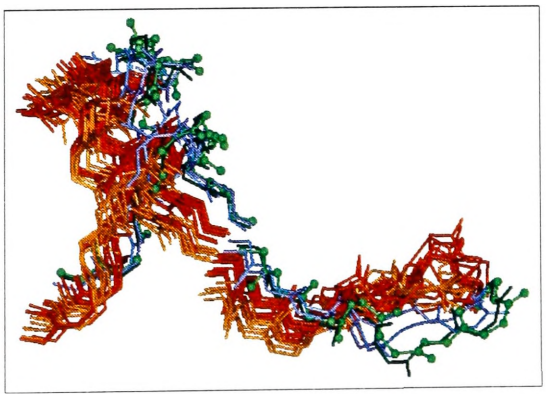


Figure 3.35 (Left, above) Two different views of the displacement of the backbone of the template grip upon substrate and inhibitor binding. 1dlo (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM (orange), and OCMS structures (red).

3.5.3.Conformational changes for Leu74-Asp76-Arg78

Of the substrate-bound and unliganded structures, only 1dlo and 1rtd had resolved sidechains for Leu74, and these were notably different from each other. In 1rtd the sidechain of Leu74 was able to make contact with the template nucleotide that is base paired with the incoming dNTP. In 1dlo, this residue, while not displaced back into the fingers, showed a very small movement upward in the plane of the fingers subdomain and further away from the template. The backbone of 2hmi, seen as a light blue structure in Figure 3.36, showed that it was much further displaced back into the fingers away from the nucleic acid binding cleft. Residues Asp76 and Asp78 are much more similar between the four DNA-bound and unliganded structures, as seen in Figure 3.33.

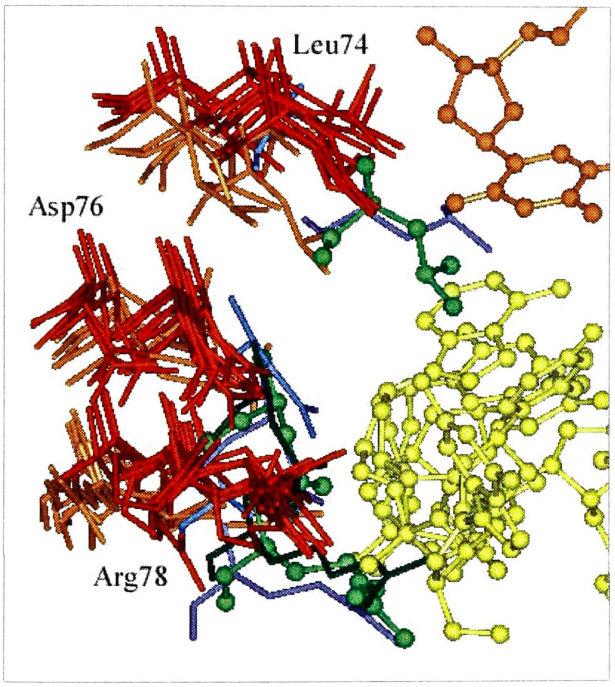


Figure 3.36 (Left) The displacement of Leu74, Asp76, and Arg78 upon substrate and inhibitor binding. The view is from above the fingers and hand. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The template strand in 1rtd is shown in yellow ball-and-stick, the incoming nucleotide is in orange.

The red and orange inhibitor-bound structures for these residues in Figure 3.36 all showed essentially the same displacement of this segment of the template grip deeper into the fingers, as if the fingers of the hand were opening up, or unclenching. Varying sidechain conformations were seen for the inhibitor-bound structures of these three residues, but the only significantly notable deviation is in the sidechain of Arg78. In most structures, including the unliganded and DNA-bound structures, the sidechain points directly upward to the top of the fingers following the plane of the subdomain. The exceptions were 1hni, 1hnv, 1tvr, and 1uwb which pointed in the opposite direction, down the plane of the fingers, toward the palm. This difference in conformation explained the suddenly high line traces of some of the inhibitor-bound structures for this residue in Figure 3.32. Leu74 in all the inhibitor-bound structures showed an additional

shifting upward in the plane of the fingers subdomain to a vertical position similar to the unliganded 1dlo structure.

Mutations of Leu74 have been shown to affect the fidelity of DNA synthesis,[162] with speculation that the residue is involved in template repositioning. An amino acid change at this position could lock the templating base tightly in place, possibly also influencing dNTP binding through interactions with the sidechains of Arg72 and Gln151.[136] There was insufficient evidence in the small changes seen between the 1rtd and 1dlo structures of Leu74 to theorise that the conformational changes were substrate induced. It may be possible that Leu74 undergoes a conformational change to bring it close enough to the *n* nucleotide about to be base-paired, to stabilise it so that it is correctly identified for base pairing. Alterations in the structure of the residue may perturb the correct stabilisation and identification of the template base, thus affecting the fidelity of replication. The inhibitor-induced conformational changes seen in Figure 3.36 would surely reposition Leu74 too far away from the *n* base of the template strand to allow it to function as speculated in its normal role.

Residue Asp76 has been postulated to interact with the template nucleotide that base pairs with the incoming dNTP, promoting mispairing by occasionally interfering with base pairing of the template nucleotide to the incoming nucleotide.[184] Mutations of Asp76 have the effect of increasing replicational accuracy, and may not have been seen in clinical isolates because increased fidelity may be detrimental to viral fitness.[185] Only small differences were seen in the line traces of Figure 3.33 between the unliganded and the DNA-bound structures for Asp76; the blue and green structures of Figure 3.36 were also quite similar. As literature appears to indicate that the role played by Asp76 in DNA replication is not vital, it is likely that the conformational displacements induced by inhibitor binding (Figure 3.36) did not have a direct affect in the inhibition of the enzyme, but were rather due to the shifts observed for other residues of the template grip, and possibly other structural motifs.

Arg78 is believed to interact directly with the phosphate backbone between *n*, the template nucleotide about to be base paired to the incoming dNTP, and *n*-1, the template nucleotide already base-paired to the last nucleotide of the primer strand,[184] as well as Asp76. Literature revealed little about possible roles it might play in template positioning and ultimately DNA replication, but they are likely to include a structural supportive role as evidenced by its likely interaction with Asp76. The results of the

DDM-based superimposition definitively showed that the binding of non-nucleoside inhibitors induced a displacement of Asp78 deeper into the fingers, and away from the nucleic acid binding cleft, as well as altering its structural relationship with Asp76. The interpretation of these results requires further experimental and mutational evidence of the role played by both Asp78 and Asp76 in the function of the template grip.

3.5.4 Conformational changes for Glu89-Val90-Glu91-Leu92

This short section of residues along the palm of the hand showed several notable differences between the unliganded and the DNA-bound structures, most especially for residues Glu91 and Leu92. Only small differences were seen between both DNA-bound structures and the two unliganded structures for Glu89 and Val90, as indicated by the black and grey line traces of Figure 3.33. Generally the DNA-bound structures were placed slightly more toward the fingers domain than the unliganded structures, as illustrated clearly by the blue structures shifted left compared to the green structures of Figure 3.37.

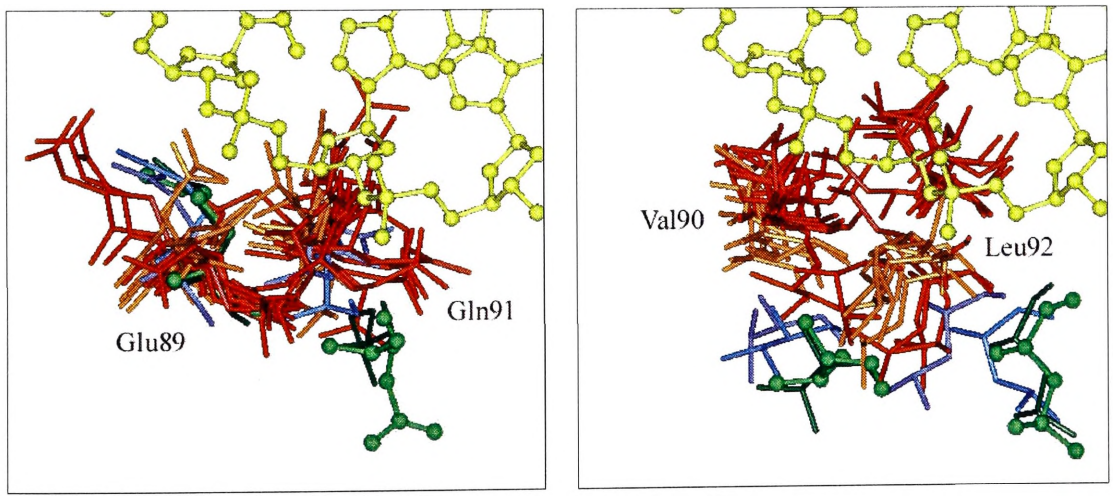


Figure 3.37 (Left) The displacement of Glu89 and Gln91 upon substrate and inhibitor binding. (Right) The displacement of Val90 and Leu92 upon substrate and inhibitor binding. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The template strand in 1rtd is shown in yellow ball-and-stick.

The sidechain of Glu89 in the unliganded and DNA-bound structures pointed in the general direction of the first finger, or toward the back of the hand, with the clustering of the blue and green structures of Figure 3.37 (left) giving an indication of how conformationally similar they were. The sidechain of Val90 in these four structures was oriented to point downward into the palm, though the substrate-bound structures

showed a displacement toward the fingers and the back of the hand. In none of the structures was Val90 within contact distance of the template strand. Both DNA-bound structures displaced Glu91 higher in the plane of the palm than the green 1dlo unliganded structure of Figure 3.37 (left) and the residue pointed toward the back of the hand and the thumb, in a downward direction. The backbone of this residue in the unliganded structure was lower in the plane of the palm, and the sidechain curved to point down into the palm, away from the template. Residue Leu92 was quite similar in the 2hmi substrate-bound structure to both unliganded structures, with the sidechains directed toward the wrist and RNase H subdomain, while in the 1rtd structure this residue was displaced notably toward the fingers subdomain (Figure 3.37, right).

In the inhibitor-bound structures, Glu89 did not show significant displacement in the backbone, but the sidechains were different to the unliganded and DNA-bound structures, all pointing toward the RNase H subdomain along the plane of the palm. The exceptions to this were the 1rt2, 1rt3, and 1rev structures which pointed upward toward the middle of the fingers (Figure 3.37, left). The sidechains in 1bqm, 1bqn, 1hni, 1hnv, and 1uwb also differed as they pointed upward toward where the template strand would lie, close enough to make contact with where its sugar-phosphate backbone would lie in the 1rtd structure. In the inhibitor-bound structures, Val90, with the exception of 1rt2 and 1rt3, always showed a displacement away from the unliganded position toward the back of the hand, or toward the PAS. This was illustrated as an upward shift in Figure 3.37 (right), as well as a vertical displacement upward from the plane of the palm. The structures of 1rt2 and 1rt3 were instead similarly placed to the unliganded and DNA-bound structures, but still showed the vertical displacement of the other inhibitor-bound structures. In Val90 this upward displacement did not extend sufficiently to force the residue into the space where the nucleic acid would lie. The sidechains of Glu91 were varied and inconsistent in their displacements in the inhibitor-bound structures, but most were displaced backward, closer toward the back of the hand and the PAS, and slightly higher in the plane of the palm. However, these displacement generally did not force Glu91 to contact the template strand, as it would lie in the catalytic 1rtd structure, because of the backward displacement toward the PAS previously mentioned. A higher displacement in the palm and a shifting left toward the fingers subdomain did cause Leu92, in most of the inhibitor-bound structures, to intrude

into DNA-binding cleft sufficiently to occupy the space where the template strand lies in the 1rtd structure.

3.5.5 Conformational changes for Gln151

This residue in 1dlo and 1rtd was similarly positioned, with the sidechain pointing toward the back of the hand in the direction of the thumb, or toward the incoming dNTP in the 1rtd structure. The only difference in conformation of the sidechain was a swivelling about the C_γ atom of about 90° (Figure 3.38). In the light blue 2hmi and dark green 1hmv structures, Gln151 was displaced slightly deeper into the fingers, toward the left of Figure 3.38, and its sidechain pointed into the fingers subdomain in the direction of the first finger. The slight differences in the conformations of the two DNA-bound structures may be due to the presence of the incoming dNTP in 1rtd, that is lacking in 2hmi, as well as the different conformations of the template strands resolved by different techniques for each structure. Whether these differences compared to the 1dlo unliganded structure were substrate induced could not be reasonably hypothesised as the structure of the dark green alternate unliganded 1hmv was so different from 1dlo.

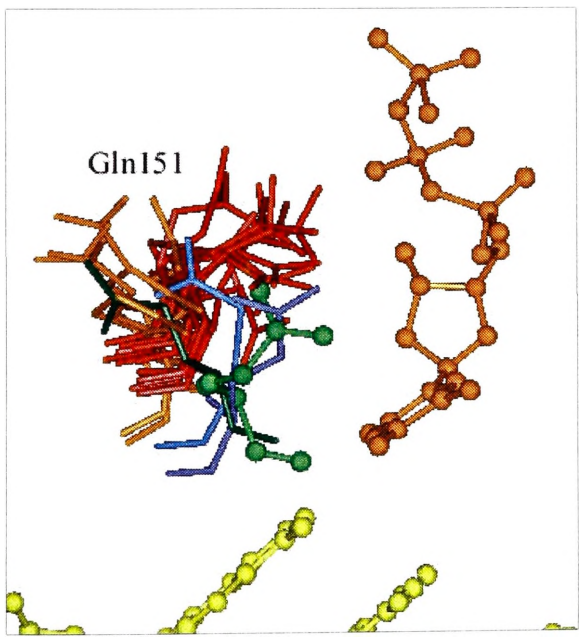


Figure 3.38 The displacement of Gln151 upon substrate and inhibitor binding. The view is from above the hand. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The template strand in 1rtd is shown in yellow ball-and-stick, the incoming nucleotide is in orange.

The inhibitor-bound structures displayed an inconsistent variety of conformational changes. Those from the CABM research group were displaced in the plane of the fingers more toward the first finger or toward the back of the hand, and had the Gln151 sidechain pointing upward to the top of the fingers, and unable to make contact with the incoming dNTP. A very slight displacement deeper into the fingers was also seen. Structures from the OCMS research group displayed a similar backbone displacement, but their sidechains generally were still directed toward the incoming dNTP, such that those

sidechains in 1klm, 1rt1, 1rt2, 1rt4, 1rt6, 1rt7, and 1rth were still within interacting distance with the dNTP.

Gln151 is a constituent of the highly conserved LPQG motif present in all reverse transcriptases, and is thought to interact directly with the 3'-OH and the β-phosphate of the dNTP.[136] It also forms a part of the flexible binding pocket for the incoming dNTP and exerts both direct and indirect influences on other residues of the dNTP-binding pocket. It interacts directly with the sidechain of Arg72, that is implicated in stabilising the transition state ternary complex before and after the phosphodiester bond formation.[161,186,187] Mutations of Gln151, thought to be unable to stabilise this sidechain, have been found to be devoid of pyrophosphorolytic activity on RNA-PBS template-primers, necessary as the release of pyrophosphate (PP_i) to complete polymerisation. Gln151 also indirectly affects the mainchain carbonyl oxygen of residue 73 that hydrogen bonds with Gln151, bridging the palm and fingers subdomain.[188] Gln151 is also hypothesised to influence the ability of the enzyme to recognise and discriminate against the sugar moieties of nucleotide substrates, due to its influence on the flexibility of the dNTP binding pocket.[188] Researchers believe Gln151 may have different and independent roles to play in reverse transcription depending on whether the template is DNA or RNA.[187,189]

The multiple and exacting roles that Gln151 is believed to play in both the formation of the phosphodiester bond and the release of pyrophosphate, the recognition of different nucleic acid templates, as well as in the stabilisation of other significant residues, would likely mean that even a slight change in the conformation and position of the residue could affect the performance of its function. While slight changes may be necessary for the dynamic and varied functions Gln151 serves, changes too great or in the wrong direction, or changes affecting flexibility may restrain the residue from all its functions. The conformational changes induced by inhibitor binding, seen in Figure 3.38, could be sufficient to exact such change on Gln151 and restrict its full functioning, and extend to influencing other residues that are indirectly dependent on the correct positioning and conformation of Gln151.

3.5.6 Activation and inhibition of the template grip

This DDM-based superimposition study has allowed quantification of the displacement of the residues of the template grip upon substrate and inhibitor binding, but in the

absence of more supporting evidence, such as thorough mutational and activity data, limited conclusions could be made about substrate or inhibitor-induced conformational changes. The comparisons between the unliganded 1do structure and the two DNA-bound structures have revealed two short sections of residues that might tentatively be considered to display substrate-induced conformational changes. Residues 73-75, of the segment of the template grip running across the middle of the fingers, in the catalytic 1rtd structure were displaced downward in the plane of the fingers subdomain, and forward toward the template strand, compared to the unliganded structure (Figure 3.34, left, top). The same residues in the binary DNA-bound structure of 2hmi showed the same displacement down the plane of the fingers, but were displaced deeper into the fingers rather than forward toward the template strand. This difference between the two DNA-bound conformations may be induced by the incoming dNTP.[136] Residues 73-75 may be part of a large closing down of the fingertips to momentarily trap the template strand and the dNTP, that is possibly held in place by its contact with Leu74[136,160]. Accordingly, 1rtd represents the 'closed' fingers state where the template strand and dNTP are positioned for optimal base-pairing and imminent polymerisation with the 3'OH of the primer terminus. Following phosphodiester bond formation the fingertips are postulated to bend back to an 'open' position to release pyrophosphate, allow template translocation and the binding of the next dNTP.[136] The 2hmi structure is most likely to be a snapshot of this 'open' fingers conformation.

Along the palm segment of the template grip residues Val90 and Glu91 showed differences between the two DNA-bound structures that were in contrast to the relative conformity seen in the neighboring residues 86-89. Both 1rtd and 2hmi were differently conformed to the unliganded 1dlo structure, but these differences did not appear sufficient to bring them within contact distance of the phosphate backbone to affect any influence on the positioning or support of the template strand (Figure 3.37). The differences between 1rtd and 2hmi were also difficult to rationalise, as no one structure appeared to have a significantly greater association with the template strand than the other. It could be speculated that the conformations of the structures play varying supportive roles to surrounding residues or secondary structures, and slight differences may be sufficient to alter the local architecture to distinguish between the two conformational states these structures are believed to represent.

The binding of non-nucleoside inhibitors induced conformational changes of the template grip that were much more distinct than those seen in substrate-bound structures. Both segments of the template grip composed of elements from the fingers subdomain showed displacements deeper into the fingers, or toward a more 'open' form using the analogy of the hand. These conformations forced the residues further away from the template strand, ending any stabilising contacts they would have had with the bases of the nucleic acid. This would most likely have an effect upon the processivity of polymerisation, but this could not be verified, as no data on how non-nucleoside inhibitors affect enzyme processivity was available. The permanence of these conformational changes can also only be speculated at, as the fingers subdomain have been noted to be flexible in their motions,[137] and on substrate binding could move to compensate the changes seen in these DDM-based superimpositions.

Part of the segment of the template grip that lies across the palm subdomain saw emphatic conformational changes in the inhibitor-bound structures that were most likely to affect the binding of the template strand in the nucleic acid binding cleft. Like the residues of the PAS on the palm, but further toward the back of the hand, this segment of the template grip showed a vertical displacement up from the plane of the palm, and back toward the back of the hand. Leu92 in most of the inhibitor-bound structures was displaced sufficiently to occupy the space of the template strand in 1rtd. Neighboring residues were also vertically displaced, and while these displacements did not have the same impression upon the template strand as did Leu92 in Figure 3.37 (right), it would be reasonable to expect that they may also have been propagated to surrounding regions of the protein. This may possibly influence other contacts the template strand could have with the palm subdomain.

The seemingly passive and supportive role that the template grip plays in the complete process of reverse transcription has meant that less progression has been made toward understanding its exact functions as a whole and the functions of its individual residues. The significance of the influence the displacements of the template grip induced by non-nucleoside inhibitor binding, seen in these DDM-based superimpositions, have on the inhibition of the enzyme can only be guessed. But it is likely to be notable, as the exact and complete process of reverse transcription requires the synchronous performance of all the components of the replication machinery, and no one component, such as the template grip can be deprecated. Further concerted and focussed research on the

biological functions of the residues of the template grip will be needed to support or challenge the results from these superimposition studies.

3.6 The RNase H active site

3.6.1 The RNase H active site and hydrolysis of RNA

The ribonuclease hybrid (RNase H) activity is the least understood and studied of all the functions of HIV-1 RT even though its role in reverse transcription is equally as important as polymerisation. The RNase H subdomain of HIV-1 RT has been crystallised in free form,[190] however this isolated form is not active. Embedded in the structure of the p66/p51 heterodimer, the RNase H subdomain is a carboxy-terminal extension of the p66 connection subdomain, and interacts predominantly with the p66 connection and the p51 thumb subdomains[191] (for the structure of the heterodimer, refer back to Figure 3.1). The RNA hydrolytic activity of this internal enzyme selectively degrades the RNA template strand when it is part of an RNA-DNA hybrid; when hydrolysis occurs on an RNA homopolymer this activity is designated RNase H*.[129] RNase H cleavage can occur if the rate of DNA polymerisation is lower than the rate of RNA hydrolysis.[192]

The HIV-1 RT RNase H domain is folded into a five-stranded β-sheet flanked by four asymmetrically arranged α-helices. The RNase H active site is located in a shallow cavity containing four acidic amino acid residues Asp443, Glu478, Asp498, and Asp549, as well as His539 (Figure 3.39). All five residues are highly conserved in retroviral and bacterial RNase H; the mutation of a single residue from this group renders the RNase H subdomain inactive.[193-195]

RNase H has at least three specific functions in the reverse transcription process[191,196] (see Section 3.1.2). First, the enzyme partially degrades the 5'end of the viral RNA template during or after (-)ssDNA synthesis, leaving intact a specific polypurine tract (PPT) to serve as an RNA primer, and facilitating the first strand transfer. Second, it excises the PPT primer from the (+) strand DNA, by hydrolysis of the phosphodiester bond at the RNA-DNA junction, for the initiation of second-strand (+) DNA synthesis. Finally, the RNase H removes the (-) strand tRNA and the (+) strand primer, enabling synthesis of full-length dsDNA to be completed.

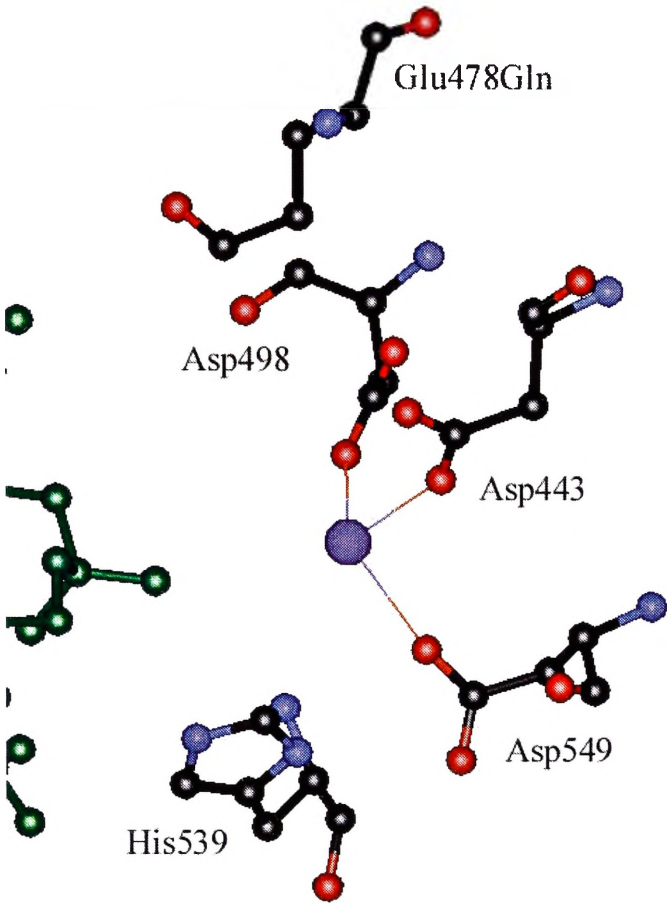


Figure 3.39 Proposed coordination complex of the RNase H active site of the catalytic 1rtd structure with possible interactions with the Mg^{2+} ion; Mg^{2+} ion (purple), template strand (dark green). The residues of the active site are coloured by atom, oxygen (red), carbon (black), and nitrogen (blue).

Two distinct modes of RNase H activity have been identified, a polymerase-dependent and a polymerase-independent mode.[128,197-199] In the polymerase–dependent mode, RNase H degrades the RNA template upon elongation of the DNA strand, and its specificity is defined, in part, by binding of the PAS to the 3'terminus of the primer strand.[200] The RNase H active site is believed to need only be positioned at a site transiently during reverse transcription to enhance the hydrolysis at the site, resulting in products of 18-20 bases in length.[196] The polymerase-independent activity of RNase H is not coupled to polymerase catalysis, and may possibly be sequence dependent.[196] It results in degradation products of down to seven bases, as it can catalyse RNA hydrolysis at many sites in the RNA up to six nucleotides from the ribonucleotide complementary to the 3'terminal deoxynucleotide of the primer. Both modes of RNase H activity are required for the completion of the enzyme's role in reverse transcription. Removal of the tRNA primer by RNase H, one nucleotide removed from the DNA-

RNA junction, leaving a single ribonucleotide 5-phosphate at the 5'terminus of the DNA genome, can be polymerase dependent[196] or independent.[201]

RNase H exhibits 3'-5' exonuclease activity in addition to endonuclease activity.[198,202] The exonuclease activity removes nucleotides one at a time from the end of the template strand, while the endonuclease activity cleave the phosphodiester bond from within the RNA strand. Studies have suggested that the RNase H domain and active site may function in two conformations; one in which both the endonuclease and exonuclease activities occur, and another in which only the endonuclease is active.[203] The conformational state of the enzyme may depend on the nature and structure of the nucleic acid substrate, and the divalent cation activator at the active site.

The contention with regard to the divalent metal ion(s) involved in the RNase H activity of HIV-1 RT has been wide-ranging and ongoing. Although, most studies appear to agree that Mg^{2+} is the preferred ion of choice, other ions have been shown to be almost as effective. The replacement of Mg^{2+} by Fe^{2+} in the RNase H active site confers the ability to cleave dsDNA. The cleavage mechanism, oxidative or enzymatic, as well as the positions of the cut can vary depending on whether DNA or RNA, respectively, is the template.[192] In some mutated enzymes that have lost natural RNase H activity, the addition of Mn^{2+} has restored activity, suggesting that the $Mg^{2+}$and Mn^{2+} occupy different active sites within the mutant enzyme. The alternative coordination and stabilisation of specific conformations of the active site that Mn^{2+} may allow, could possibly compensate for the mutation.[195] The question of which divalent ion is naturally used by RNase H is still not clarified with the available crystal structures. The ternary catalytic structure bound with dsDNA (1rtd), and an inhibitor-bound structure (1rev), showed only one Mg^{2+} ion,[136,138] although the ion in the latter structure was not fully identified, and only modelled on the assumption that it was a magnesium ion. In Figure 3.39 the structure of the RNase H active site in 1rtd is shown with respect to the magnesium ion and the template strand, which were separated by only 3.72 Å. The contact between Asp443 and the Mg^{2+} ion was verified in the crystal structure,[136] while the contacts to Asp498 and Asp549, at 3.16 Å and 2.53 Å, respectively, are only postulated. Glu478 was mutated to a glutamine residue to inactivate the RNase H activity in crystallising the 1rtd structure. Mn^{2+}-doping studies, as well as a crystal structure of the free RNase H subdomain, have revealed two bound Mn^{2+} ions, separated by approximately 4 Å.[190,204] This alludes to a hypothesis that there may exist

two distinct and mutually exclusive Mg^{2+} binding sites; depending on the reaction that is catalysed, with only one metal ion site populated for any specific reaction.[204]

The catalytic mechanism of RNase H hydrolysis of the phosphodiester bond has not yet been elucidated. The metal ions have been speculated to play a role in the stabilisation of the negatively charged nucleophile (OH^-) and the transfer of the negative charge from the nucleophile to the phosphate.[191] Two alternative catalytic mechanisms have been proposed to satisfy the contentious issue of how many ions are involved.[205] One is a two-metal ion mechanism, in which the hydrolysis of the P-O3' bond is promoted by two metal ions, such that the first metal ion activates an attacking water molecule, and the second metal ion stabilises the transient pentavalent phosphorus intermediate and facilitates the release of the 3'oxyanion. The other is a general acid-base mechanism, in which His539 activates an attacking water molecule as a general base and a single metal ion stabilises a transient intermediate by forming an outer sphere complex.

The PAS and the RNase H active sites can contact an RNA-DNA duplex simultaneously; the contact points lie approximately 18 base pairs apart on the double-stranded nucleic acid, though this number has been the source of much debate.[132,136,196,197,199,202,206] In the catalytic 1rtd structure there are 17 base pairs between the two active sites, and 17-18 base pairs for the binary 2hmi structure. The number of base pairs between the catalytic sites is dependent on the state of the enzyme as the protein, the nucleic acid, or both, are expected to undergo conformational changes as part of the polymerase and RNase H reactions.

The RNase H subdomain and its activity are not autonomous; mutations in the primer grip,[171,172,207] finger, and thumb[208] subdomains can alter the positioning of the nucleic acid substrates and decrease RNase H activity. The binding of the NNRTI nevirapine has been shown to cause sufficient distortion of the enzyme to affect the specificity of RNase H cleavage.[209] The lack of RNase H activity data for NNRTIs has hampered efforts to further correlate and corroborate the hydrolytic inhibition effects of nevirapine, and this deficiency needs to be addressed for further study. Investigation of the substrate and inhibitor-induced conformational displacement of the active site RNase H residues can form a basis of quantitative evidence of long range conformational changes affecting all of RT's enzymatic activities. Along with further experimental evidence, this data may be able to fully assess and assign the functions of

the residues and explain how substrate and inhibitor binding affect their roles in polymerisation and hydrolysis.

3.6.2 Synopsis of the displacement of the RNase H active site

The results of the superimposition showed clearly the small differences in positioning and conformation induced by substrate binding, and the larger displacements that were induced by inhibitor binding (Figures 3.40 and 3.41). The residues of the RNase H active site are listed in Appendix 7.6, and displacement data is tabulated in Appendix 7.12.

The comparison of 1dlo with the second unliganded structure 1hmv, on the whole, was favourable with low RMSD values for the superimposition. The only deviation seen between the unliganded structures was in residue His539, where the light grey square line trace of Figure 3.41 and 3.42 showed an average displacement of 1.98 Å, compared to 0.76 Å for the remaining four residues.

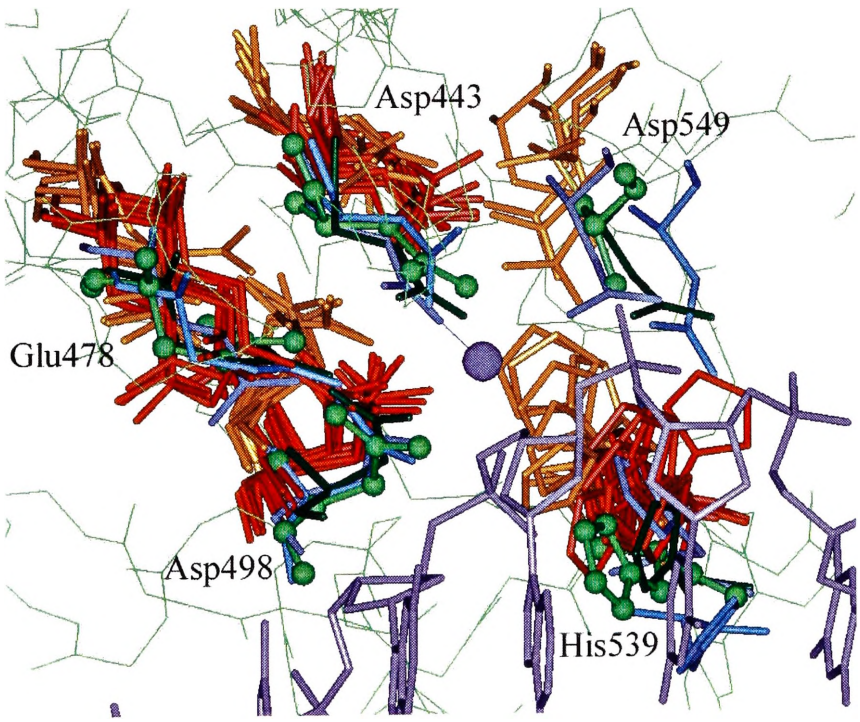


Figure 3.40 The residues of the RNase H active site, upon substrate and inhibitor binding. The connection and thumb subdomains are to the left of the illustration. The backbone of 1dlo is shown as a green line trace, and the template strand and Mg^{2+} ion in 1rtd is shown as a purple stick structure and purple sphere respectively. Red and orange active sites represent inhibitor-bound structures, light and dark blue active sites represent DNA-bound structures. This view is the same used in the individual residue analysis, with details of the conformational changes in the text.

The binary and ternary RT-DNA structures showed similar conformational changes, with the differences being predominantly in the sidechain displacements which showed greater variation. The conformation and placement of His539 in both DNA-bound structures was also quite different from 1dlo, resulting in high black and grey line traces for this residue in Figures 3.41 and 3.42, that were not easily discernible from the multi-coloured line traces of the inhibitor-bound structures. Like the residues of the PAS, the RNase H residues of the catalytic 1rtd structure showed a greater similarity to the unliganded 1dlo structure, than the binary 2hmi structure, particularly in terms of the average backbone displacement. The inhibitor-bound structures all showed considerable shifting of the RNase H active site residues, with displacements consistent in magnitude and direction, as well as conformation. Clustering effects due to the origins of group crystal structures were most notably seen in Asp498 in Figure 3.41, where the two groups of multi-coloured line traces were separated by an average difference of about 1.66 Å. The average displacement values of Table 3.7 for the two groups of crystal structures, in all separate components, consistently varied by approximately 1.0 Å. A clear delineation of the different conformational displacements could be seen between unliganded or DNA-bound structures and inhibitor-bound structures for all residues save His539, and the sidechain of Glu478 (Figure 3.41).

average displacement (Å)		DNA-bound structures		inhibitor-bound structures		
	1hmv	ternary	binary	total	CABM	OCMS
entire residue	1.07	1.21	1.36	3.02	3.72	2.69
backbone	0.91	0.85	1.03	2.90	3.60	2.57
sidechain	1.22	1.60	1.69	3.12	3.83	2.80

Table 3.7 Summary of the average displacement between the 1dlo unliganded structure and the second unliganded (1hmv), substrate-bound (1rtd-ternary and 2hmi-binary), and inhibitor-bound structures for the entire RNase H active site, and only the backbone or sidechain component. The three representative average displacements were calculated for all 19 inhibitor bound structures (total), and based on the research groups that derived the crystal structures; six from the CABM research group, and 13 from OCMS research group.

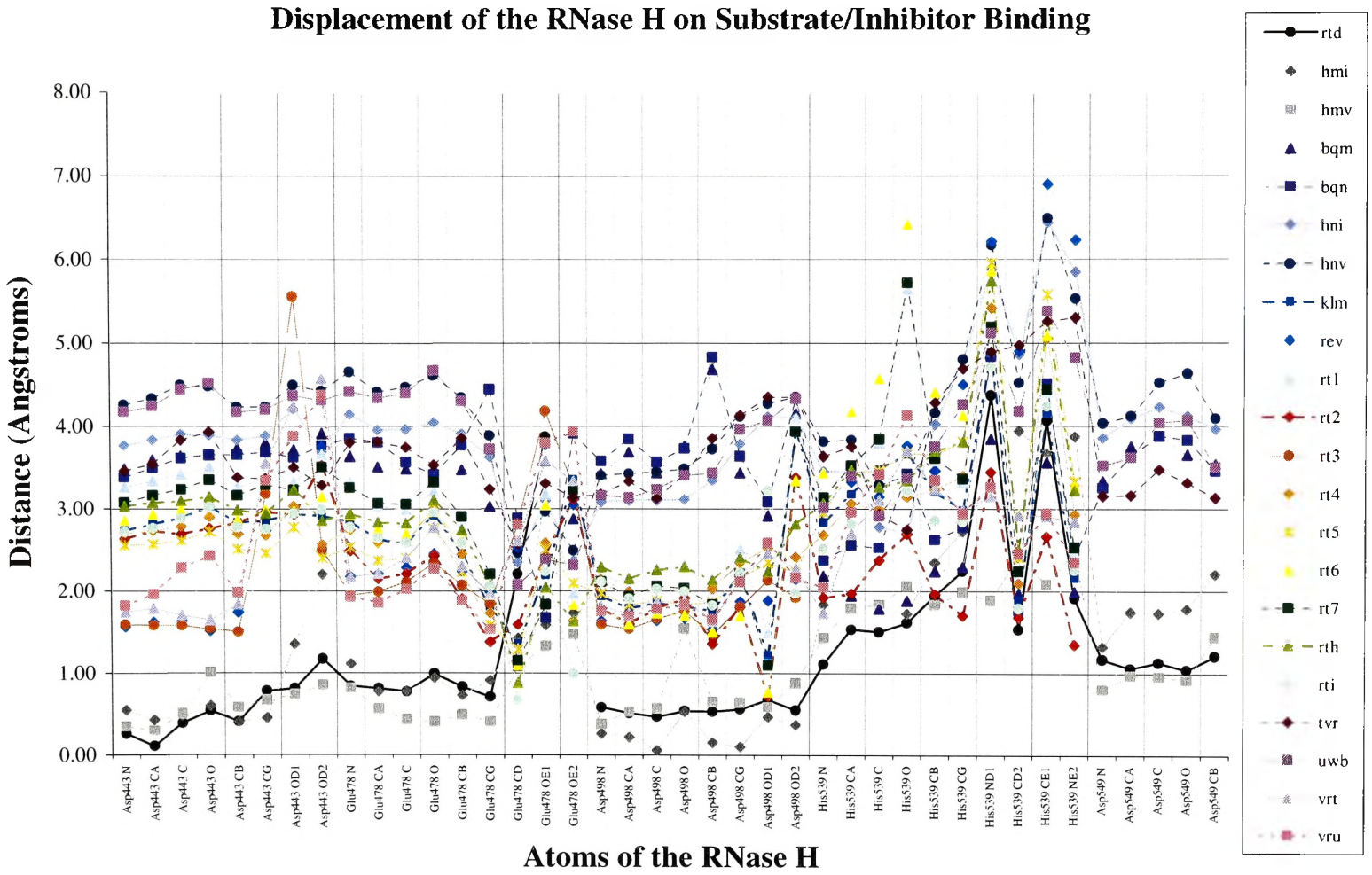


Figure 3.41 Graphical representation of the displacement between the unliganded (1dlo) structure and all liganded RT structures for the RNase H.

3.6.2.1 Substrate induced conformational changes of the RNase H active site

The RNase H active site was situated in a shallow groove, at the mouth of which the template strand runs (Figure 3.40). The roof of this groove was formed by a β-strand flanked by two α-helices; Glu478, Asp443 and Asp549 lie side-by-side across the roof of the groove. A β-strand and a connecting loop containing Asp498 and His539 respectively form the floor of the groove. In the 1rtd structure a Mg ion was coordinated to the $O_{\delta 2}$ of Asp443, and was approximately situated in the middle between Asp443, Asp549, Asp498, and His539, at the centre of the groove.

Residues Glu478 and Asp443, at the top of the groove, in both DNA-bound crystal structures did not show any significant displacement relative to the 1dlo unliganded or 1hmv structure. This is clearly demonstrated by the overlapping blue and green structures in Figure 3.40, and the thin dark sticks and pale ball-and-stick structure of Figure 3.43 (left) compared to the thick sticks of 1dlo. The slight exceptions seen in the black and dark grey line traces of Figure 3.42 were in the sidechain of 2hmi's Asp443, due to a torsion about the C_β atom changing the conformation of the aspartate fork, and the sidechain of Glu478Gln in 1rtd. The displacement in this sidechain of 3.87 Å may be due to the deliberate mutation to inactivate the RNase H activity. Relative to the uniformity of Glu478 and Asp443, the small average displacements of Asp549 in 2hmi and 1rtd, of 2.71 Å and 2.18 Å respectively, attracted some slight attention for their disparity. In 2hmi, the displacement was represented by the dark thin stick structure at the bottom of Figure 3.43 (left), shifted relative to the dark thick structure.

Asp498 at the bottom of the groove in the DNA-bound structures was identically orientated and positioned to the unliganded structure, as evidenced by the line traces of Figure 3.42, showing an average displacement of 0.55 Å and 0.27 Å for 2hmi and 1rtd respectively. His539 conversely showed a tangle of line traces for this residue, that actually only appeared notable because of the small scale of displacements seen in the other residues of the RNase H active site (Figure 3.42). The sidechains of this structure at the bottom of Figure 3.43 (left) show the disparity in conformation between the unliganded and DNA-bound structures. This has been proposed to be a catalytic residue, so the changes observed upon substrate binding could be significant.

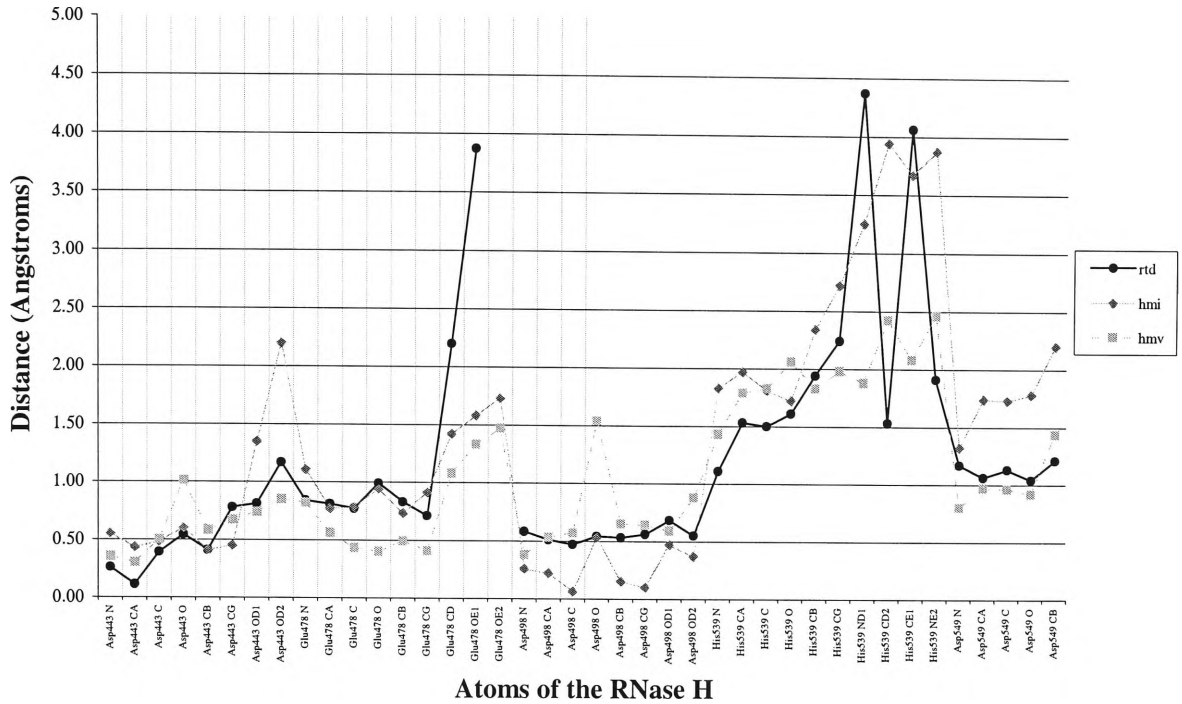


Figure 3.42 Graphical representation of the displacement between the unliganded (1dlo) structure, and the second unliganded (1hmv) and DNA-bound structures (1rtd and 2hmi) for the RNase H active site.

3.6.2.2 Inhibitor induced conformational changes of the RNase H active site

All of the residues of the roof of the groove in the inhibitor-bound structures showed a small common displacement toward the thumb, and a much more significant displacement inward into the groove and away from where the template strand would lie at the mouth of the groove. This is seen as shifts toward the top of Figure 3.43 (right) and shifts toward the right side of the illustration. As such, the groove containing the active site appeared to be displaced out, away from the nucleic acid binding cleft. This could be better described as an elaboration of the opening of the hand-like structure of the enzyme seen in the fingers and thumb domains in the studies of the primer and template grips. The crystal structures of the CABM group generally showed greater displacements than those of the OCMS group, as seen in the clustering and displacement of the orange coloured structures of Figure 3.40 toward the top of the illustration, compared to the red coloured structures.

Asp498 and His539, at the bottom of the groove, in the inhibitor-bound structures showed a similar consistent displacement deeper into the groove, and a small displacement toward the thumb (Figure 3.43, right). A small elevation was seen in the inhibitor-bound structures for these residues, as if the groove had been slightly compressed with the floor of the groove elevating toward the roof. Residue His539, on

the floor of the groove, did not appear to be moved as consistently as those at the roof of the groove or Asp498, as clearly seen in the erratic-like displacements measured for His539 in Figure 3.41. The inhibitor-bound structures represented by 1hni and 1vru in Figure 3.43 (right) show clearly the increased distance between the RNase H active site upon inhibitor-binding and the template strand that must be degraded for the successful completion of reverse transcription. The altered relationship between the active site residues could make it impossible to correctly coordinate the Mg^{2+} ion for hydrolysis, even if other motifs in the enzyme were able to compensate for inhibitor binding and adapt or readjust the positioning of the template strand closer to the active site residues.

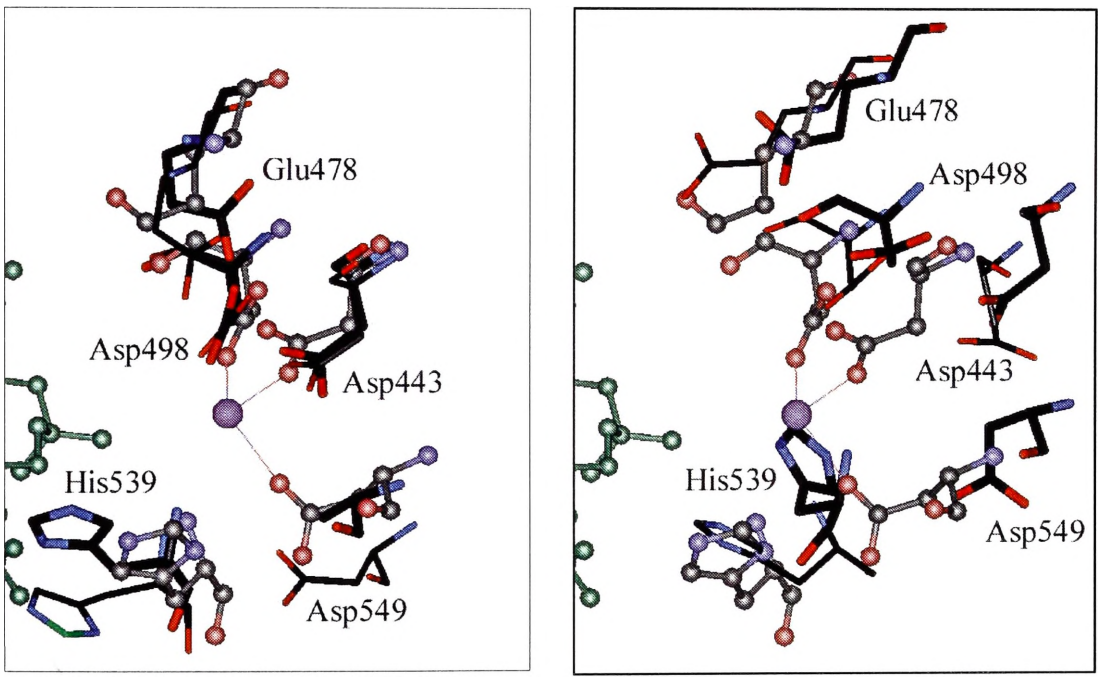


Figure 3.43 (Left) The conformational differences between the unliganded structure 1dlo (thick dark sticks) and the binary DNA structure 2hmi (thin dark sticks), seen relative to the proposed catalytic coordination complex of RNase H hydrolysis in 1rtd shown in pale ball-and-stick (see Figure 3.39 for full legend details). (Right) Inhibitor-bound structures of 1hni (thick dark sticks) from the CABM group and 1vru (thin dark sticks) from the OCMS group, also seen relative to 1rtd.

3.6.3 Displacement of aspartic acid 443

Asp443 is situated in the middle of the roof of the RNase H groove, and showed virtually identical conformations and negligible displacements between all four unliganded and DNA-bound structures. The Y-shaped carboxylate fork of the Asp443 structures in both unliganded structures and the ternary DNA-bound structures pointed toward the Mg^{2+} ion at the centre of the groove, to which it is coordinated in the 1rtd structure. The carboxylate fork of the binary DNA structure (2hmi) showed a swivel of

approximately 90° at the C_γ atom, which would be unlikely to affect its coordination to the Mg^{2+} ion.

The similarity of conformation and position in Asp443 between the unliganded and substrate-bound structures lend support to the supposition that this residue is not affected by the binding of the nucleic substrate, and may not require conformational changes to become catalytically active. This superposition, for Asp443 and other residues of the RNase H active site, can only be tentative at the moment, as the nucleic substrate in the 2hmi and 1rtd structures was dsDNA, which is not the natural substrate of the RNase H active site.

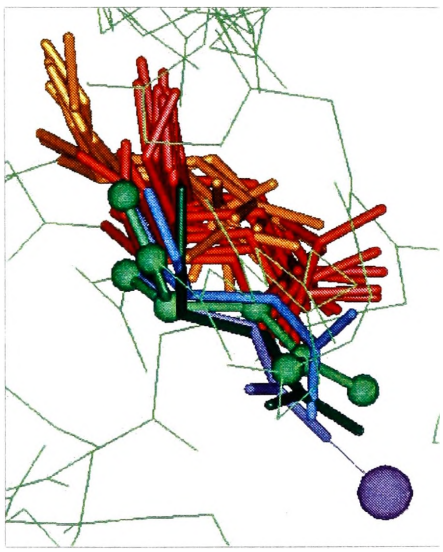


Figure 3.44 The displacement of Asp443 upon substrate and inhibitor binding. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The backbone of 1dlo is shown as a green line trace, and the Mg^{2+} ion in 1rtd is shown as a purple sphere. The thumb subdomain is to the left, as can be more clearly seen in Figure 3.40, which is in the same orientation.

The inhibitor-bound structures showed a distinct and uniform displacement deeper into the groove, in addition to a smaller displacement out, away from the thumb, as depicted by shiftings of the orange and red coloured structures of Figure 3.44 to the top of the illustration, and toward the right hand side, respectively. The average displacement of this residue in the inhibitor-bound structures was 3.15 Å compared to 0.66 Å for the unliganded and DNA-bound structures. The conformation of the carboxylate forks were generally unchanged, with the exception of 1rev, 1rt3, 1vrt, and 1vru from the OCMS group which swivelled at the C_β atom to point down to the floor of the groove. Swivelling about this atom, unlike that seen in the C_γ atom of 2hmi, further increased the displacement away from the Mg^{2+} ion, and it would be unlikely that the carboxylate fork of any of the inhibitor-bound structures could coordinate to the Mg^{2+} ion if it was similarly placed as seen in the 1rtd structure. The structures of the OCMS research group tended to be more modest in their displacements than those of the CABM group, with an average difference of 1.08 Å.

3.6.4 Displacement of glutamic acid 478

This residue is situated at the roof of the groove closest to the thumb subdomain, and is composed of a glutamic acid residue in all of the structures except for the 1rtd ternary DNA-bound structure in which a glutamine was substituted instead to inactivate the enzyme for crystallisation purposes. The dark green 1hmv and light blue 2hmi structures of Figure 3.45 were extremely similar in orientation and position; their carboxylate forks pointing toward Asp443 and the Mg^{2+} ion in 1rtd, just as in the green ball-and-stick 1dlo unliganded structure. The only small difference in the DNA-bound structures appeared in the sidechain of 1rtd that pointed out into the groove toward the template strand due to a swivel at the C_γ atom. This displacement in1rtd, seen clearly in the black circle line trace of Figure 3.42, could be explained by the mutation to glutamine in the 1rtd structure. With this small exception, the DNA-bound structures appeared not to display any substrate induced changes from the unliganded state that could be postulated to be catalytically active conformations.

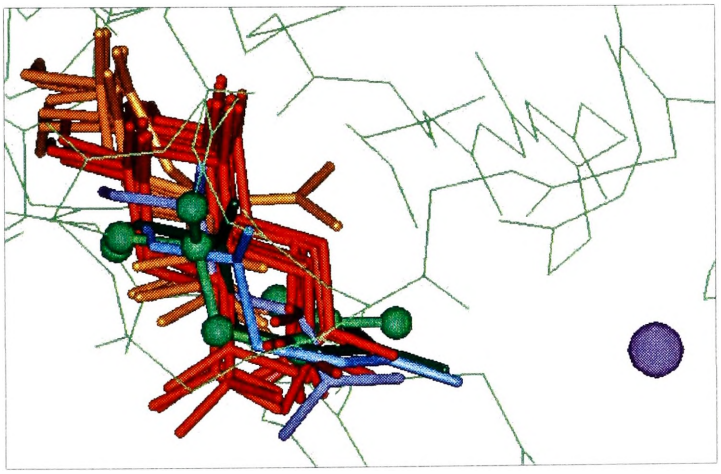


Figure 3.45 The displacement of Glu478 upon substrate and inhibitor binding. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The backbone of 1dlo is shown as a green line trace, and the Mg^{2+} ion in 1rtd is shown as a purple sphere. The thumb subdomain is to the left, as can be more clearly seen in Figure 3.40, which is in the same orientation.

The conformational displacements of Glu478 in the inhibitor-bound structures showed distinct and uniform movement deeper into the groove, with smaller displacements out, away from the thumb, similarly to Asp443. With the exception of 1bqn (orange Y-shaped structure in the middle of Figure 3.45) whose side chain is pointed toward Asp443, all the liganded structures pointed out into the groove like 1rtd. The sidechains of the red coloured OCMS structures of Figure 3.45 appear not to be as far displaced backward into the groove compared to the orange coloured structures in the top left hand corner of the illustration. However, they still had a noteworthy average displacement of 2.42 Å from the green ball-and-stick unliganded structure.

3.6.5 Displacement of aspartic acid 498

Asp498 is situated on the floor of the groove, closest toward the thumb. As for Asp443, negligible differences could be found in the orientations and positions of the four unliganded and DNA-bound structures. The blue and green structures of Figure 3.46 show Asp498 in these structures almost as one structure. All of their carboxylate forks pointed upward into the groove toward the Mg^{2+} ion without even notable swivelling about the C_β atom. Once more, the DDM-based superimposition revealed little evidence that the binding of the DNA substrates in 1rtd and 2hmi had any affect on the conformation of Asp498, and no postulations could made concerning catalytically active conformations that the residue might adopt.

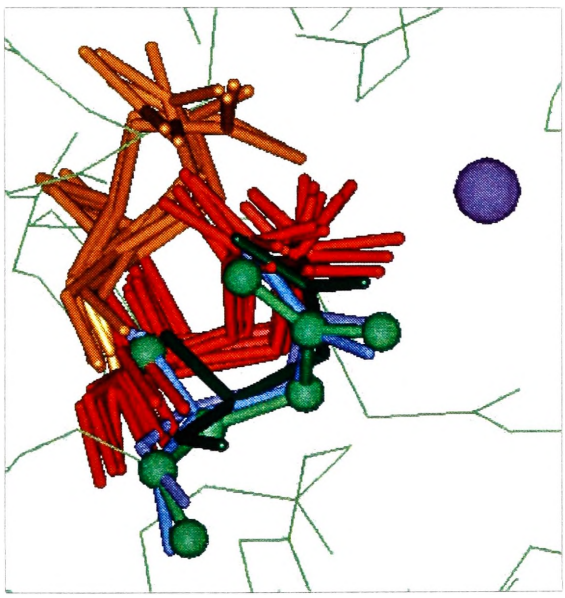


Figure 3.46 The displacement of Asp493 upon substrate and inhibitor binding. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The backbone of 1dlo is shown as a green line trace, and the Mg^{2+} ion in 1rtd is shown as a purple sphere. The thumb subdomain is to the left, as can be more clearly seen in Figure 3.40, which is in the same orientation.

Following in the same trend previously seen, the inhibitor-bound structures in Figure 3.46 showed distinct and singularly consistent displacements deeper into the groove, with smaller displacements out, away from the thumb. The conformation and orientation of this residue in the inhibitor-bound structures was comparable with the unliganded and DNA-bound structures, pointing toward the roof of the groove. Asp498 provided the most manifest example of the clustering of the residues in crystal structures from the two different research groups. Those orange coloured structures from the CABM group showed the greatest displacement deeper into the groove, as well as a small elevation upward from the floor of the groove, depicted as shifts toward the top and the left hand side of Figure 3.46. The red coloured OCMS structures were displaced, on average, 1.66 Å less than the CABM structures, but still displayed a notable displacement of 2.02 Å from the 1dlo unliganded structure.

3.6.6 Displacement of histidine 539

This residue situated on the floor of the groove furthest away from the thumb, appeared to be the most inconsistent amongst all the crystal structures, and this was reflected in the displacements measured and the erratic line traces of Figures 3.41 and 3.42.

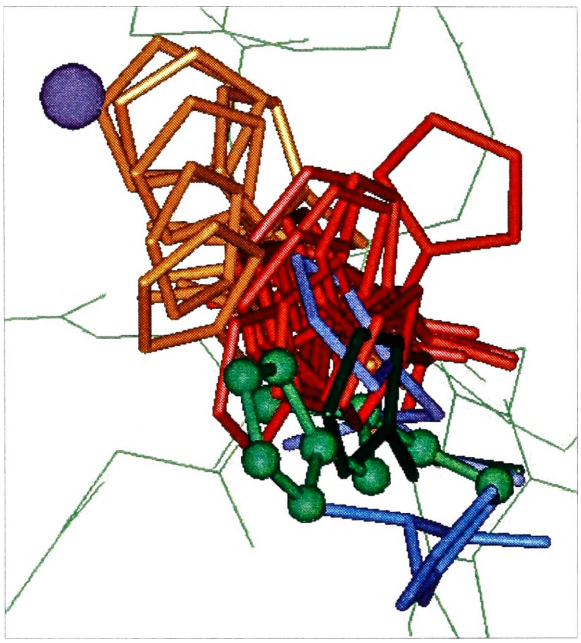


Figure 3.47 The displacement of His539 upon substrate and inhibitor binding. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The backbone of 1dlo is shown as a green line trace, and the Mg^{2+} ion in 1rtd is shown as a purple sphere. The thumb subdomain is to the left, as can be more clearly seen in Figure 3.40, which is in the same orientation.

His539 in the dark green 1hmv structure of Figure 3.47 was similarly orientated to that in 1dlo, except for a small downward depression into the floor of the groove and a displacement away from the thumb, or toward the right hand side of the illustration. The sidechain, that in 1dlo is orientated to point toward the fingers with the plane of the histidine ring approximately facing the palm subdomain, instead has its plane facing the connection subdomain. In the dark blue 1rtd, His539 showed a small displacement deeper into the groove than 1dlo. The backbone of this residue was slightly depressed down, but a change in the dihedral angle (O, C, C_α, C_β), from – 35 ° in 1dlo to – 4 ° in 1rtd, resulted in little change to the relative sidechain elevation compared to 1dlo. The sidechain in 1rtd was orientated to point more toward the thumb, while actually displaced slightly further away from the thumb, with the plane of the ring in a similar orientation to 1dlo. This all resulted in the residue in 1rtd being approximately 4.0 Å from the Mg^{2+} ion, closer than 1dlo, but not quite within contact distance. Both 1dlo and 1rtd were 2.77 Å and 3.64 Å respectively from the sugar phosphate backbone of the template strand in 1rtd. In the binary DNA-bound structure, His539 showed a small displacement out of the groove and further from the thumb, and had a slight depression downward into the floor of the groove. The sidechain was quite different from the other structures, pointing to some point over the wrist, with the plane

of the ring facing the connection subdomain, and distinctly out of contact range with both the Mg^{2+} ion and the template strand in 1rtd.

The differences between the two DNA-bound structures, and also compared to the two unliganded structures could be conformationally induced by the binding of the DNA substrate. The differences between the light and dark blue structures could be postulated to be a conformational change of the residue from a substrate-bound inactive form, to one poised to interact with the magnesium ion and the template strand for hydrolysis, though no further evidence was available to support such a postulation. The lack of greater similarity between 1hmv and 1dlo did not provide strong support for any comparisons of the DNA-bound structures with an unliganded structure.

Although the inhibitor-bound crystal structures of this residue did not show the tight clustering and uniformity typically seen in other residues, a general displacement deeper into the groove, or toward the top of Figure 3.47, was observed. With the exception of 1bqm and 1bqn, the residues also showed a displacement out, away from the thumb. The sidechains of 1bqm and 1bqn, unaffected by the gross residue displacements seen in the other inhibitor-bound structures, pointed toward the template strand in 1rtd, with a difference in distance to the sugar-phosphate backbone of less than 1.0 Å. His539 in the inhibitor-bound structures was slightly elevated from the floor of the groove, such that the sidechains of this residue in the orange coloured 1hni and 1hnv structures virtually displaced the site occupied by the Mg ion in the catalytic 1rtd structure. The greatest variation between the crystal structures was seen in the sidechains of this residue. Generally the sidechains were directed toward the fingers and thumb subdomains, with a variety of different swivelling motions of the plane of the ring, though most of them were still within contact distances of less than 4.0 Å from the sugar phosphate backbone. It is unknown what possible role His539 may play in coordinating the backbone of the template strand for hydrolysis. These DDM-based superimposition results indicate that it may be not as greatly affected by inhibitor-binding as other roles this residue may have in synchronization with the other residues of the RNase H active site.

3.6.7 Displacement of aspartic acid 549

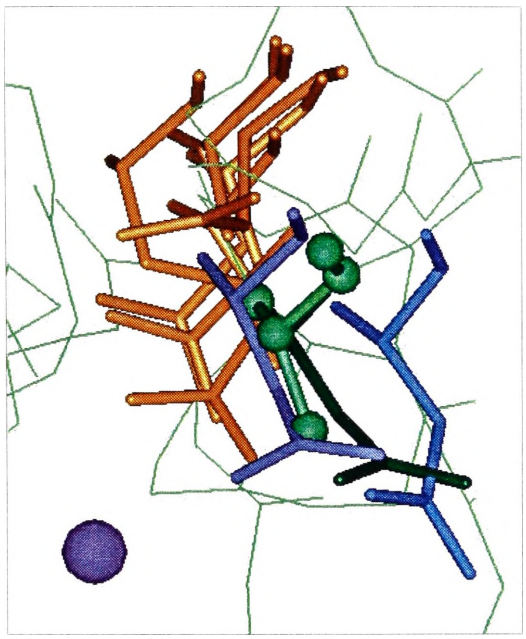


Figure 3.48 The displacement of Asp443 upon substrate and inhibitor binding. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange). The backbone of 1dlo is shown as a green line trace, and the Mg^{2+} ion in 1rtd is shown as a purple sphere. The thumb subdomain is to the left, as can be more clearly seen in Figure 3.40, which is in the same orientation.

Asp549 is situated at the roof of the groove, furthest from the thumb, and the absence of the carboxylate sidechain in the 1dlo unliganded structure limited the comparisons that could be made with this residue. Despite this, the backbones of 1rtd and 1hmv were similar in orientation, though they were both displaced slightly deeper into the groove compared to 1dlo, seen as a slight shifting toward the top of Figure 3.48. The carboxylate forks of 1rtd and 1hmv were directed downward into the groove toward His539, with the their sidechains 2.53 Å and 3.95 Å respectively from the purple Mg^{2+} ion in 1rtd. The binary 1hmi structure was similarly conformed to the unliganded and DNA-bound structures, but show a very small displacement out, further away from the thumb, and the other residues on the roof of the groove, and 4.30 Å away from the Mg^{2+} ion. This difference between the 1rtd and 2hmi structures may be a conformation change from a DNA-bound inactive form to a catalytically-poised form able to coordinate to the necessary Mg^{2+} ion.

The number of inhibitor-bound structures with which to compare this residue was less than previously used, as all crystal structures from the OCMS group did not have this residue resolved. Those inhibitor-bound structures available showed the distinct and uniform displacements deeper into the groove seen by all the residues of the RNase H active site, as well as a very small elevation deeper into the roof of the groove. The sidechains of 1hnv, 1tvr and 1uwb pointed downward into the groove, or toward the bottom of Figure 3.48, and angled slightly more toward the back of the groove than toward His539. The 1hni sidechain pointed downward into the groove, but in the

opposite direction to the connection subdomain; the sidechain of 1bqm pointed toward the fingers subdomain due to a torsion of the angle about C_β.

Although the lack of data for Asp549 made speculation tentative, the consistent displacements seen of the structures available, and their difference from the black and grey line traces of Figure 3.41 could cautiously provide a basis to postulate that Asp549 was subject to inhibitor-induced changes. These induced changes, horizontally and vertically deeper into the groove, had the effect of increasing the distance of the residue to the Mg^{2+} ion out of coordination range, and possibly conforming Asp549 into an inactive conformation.

3.6.8 Activation and inhibition of the RNase H active site

Despite the significant role it plays in the complete reverse transcription of viral RNA, no published study has been made of the conformational changes of the RNase H active site upon the binding of nucleic acid substrates or inhibitors in the NNIBP. The results of this DDM-based superimposition study have shown that the binding of NNRTIs caused singularly consistent and significant displacement of the residues of the RNase H active site. In the absence of any hydrolysis activity data, for the inhibitors in the inhibitor-bound crystal structures, and more detailed postulations of mechanism of RNase H activity, only conjectural observations could be made of these results. Based on these initial and promising findings there needs to be serious consideration to include RNase activity data when testing and studying NNRTI inhibition.

Unlike the previously studied PAS, and primer and template grips, the substrate induced conformational changes of the RNase H were difficult to interpret, and scarcely any rationalisations could be made for certain displacements. The unliganded and DNA-bound structures were all remarkably similar in conformation and placement, with the exceptions of His539 and Asp549. His539 has been suggested to interact with the sugar-phosphate backbone of the template strand in the duplex region of the template-primer.[206] The sidechain of Asp549 is postulated to be essential for coordinating the second Mn^{2+} ion in the two-metal ion mechanism,[191] as well as possibly coordinating the Mg^{2+} ion as seen in the 1rtd structure. The difference between the structures in the binary and ternary DNA-bound structures may be due to the differing lengths of the dsDNA used in each structure, and/or the presence of a Mg^{2+} ion in the ternary structure. While the template strand in 2hmi lay approximately 8 Å away from the

RNase H active site, and was not in a proper position for cleavage, the 1rtd structure had the template strand in a much more favourable orientation. It is possible that this template in 1rtd could be cleaved, if not for the glutamine mutation at Glu478. However, one needs to remember that dsDNA is not the natural substrate for RNase H, and that the conformations observed in these studies might be different for other nucleic acids.

The inhibitor-bound structures showed clearly that NNRTI binding caused a slight compression and withdrawal away from the template strand of the shallow groove containing the residues of the RNase H active site. This conformational change further reiterated the analogy of the opening of the hand. Even though the RNase H groove underwent a slight compression, as if this local motif was closing, the RNase H active site as a whole showed a displacement extending it out away from the template strand. Individual residues were displaced at a greater distance from each other, the template strand, and the Mg^{2+} ion, making catalytic hydrolysis seem unlikely. Much of the focus of the efficiency and specificity of RNase H activity has been on the appropriate positioning of the nucleic acid by the PAS,[172] and the various pitches, tilts, groove widths, and forms adopted by nucleic acid hybrids.[191,208] These factors are thought to induce the RNase H into the appropriate conformations for activity. The results from these DDM-based studies indicated that in the absence of nucleic acids, long range conformational changes of the RNase H active site could be induced by NNRTI binding. This provided preliminary evidence that structural motifs in the enzyme remote from the RNase H subdomain could play dominant roles in the conformational changes of RNase H, and possibly the nucleic acid as well. These conformational changes could be required for the recognition of correct substrates and complete hydrolysis of the phosphodiester bond, roles that may not be limited to the structure of the nucleic acid and its interactions with active site residues. This is not to say that distinctions in different nucleic acid structure could not be communicated to the enzyme to exert conformational changes. An even more confounding possibility exists that the nucleic acid structure influences such motifs as the PAS, primer, and template grips, which in turn induce conformational changes of the RNase H, and perhaps the nucleic acid itself; a vicious circle of dependency.

Two of the proposed functions of the RNase H involved interaction with the $tRNA^{Lys3}$ primer; the initial cleavage at the 5'end of the viral RNA template to facilitate the first

strand transfer, and final removal of the $tRNA^{Lys3}$ primer from the synthesised (+) strand DNA. This demands that the RNase H active site be able to differentiate between the viral RNA-$tRNA^{Lys3}$ initiation complex and the later DNA-$tRNA^{Lys3}$ hybrid. This differentiation has been seen in the PAS, where studies into the three-dimensional structure and interactions of the viral RNA-$tRNA^{Lys3}$ complex and $RT^{176,210-212}$ have suggested that conformational changes in both are required to prevent steric clashes, and provide a perfect fit of primer-template and enzyme for reverse transcription. In addition, it has been observed that RT is able to recognise the difference between RNA-$tRNA^{Lys3}$ and DNA-$tRNA^{Lys3\ 212,213}$ when extending the primer, possibly adopting different orientations of binding to allow extension. Could this ability be extrapolated to the RNase H? Both are active sites involved in enzymatic catalysis with similar divalent ions and acidic residues, both involve precise and intimate interactions with both RNA and DNA, both work in tandem and independently to reverse transcribe viral RNA. Although this question could not be answered with the present data available, it poses an intriguing and possibly fortuitous site for potential rational drug design.

3.7 The non-nucleoside inhibitor binding pocket

3.7.1 Inhibitor binding

The binding of inhibitors at the non-nucleoside inhibitor binding pocket (NNIBP) has been at the heart of the DDM-based superimposition studies. Conformational changes induced at the local level in this allosteric pocket were seen to influence the neighboring PAS and primer grip, as well as remote structural motifs such as the template grip and the RNase H active site (Sections 3.3 - 3.6). The logical conclusion to these studies was the analysis of the pocket itself, which, in the absence of an inhibitor, exists as a mere surface depression surrounded by residues Leu100, Lys101, Lys103, Val 179, Tyr181 and Tyr188 of p66, and Glu138 of p51.[30]

The NNIBP is composed primarily of amino acid residues from the β5b-β6 loop (Leu100, Lys101, Lys103), β6 (Val106), the β9-β10 hairpin (Val179, Tyr181, Tyr188, Gly190), the β12-β13 hairpin (Phe227, Trp229, Leu234, His235, Pro236), β15 (Tyr318) and the β7-β8 loop of p51 (Glu138). The structural definition of the pocket can vary amongst the literature. Most of the amino acid residues that form the binding pocket are hydrophobic and five of them have aromatic side chains. The only hydrophilic residues

around the pocket (Lys101 and Lys103 of p66 and Glu138 of p51) are located at the putative entrance.[134]

Non-nucleoside inhibitors are thought to insert between two hydrophobic surfaces of the pocket which have complementary shapes, with electrostatic interactions (charge and hydrogen bond) contributing to the final strength of binding and assisting in the orientation of the compound.[133] Specificity of binding is believed to come from the interactions with bulky hydrophobic residues, in particular from a number of ring-stacking interactions.[31,214] The intrinsic flexibility of RT, expressed here as a plasticity in the structure of the protein surrounding the bound inhibitors, allows unfavorable contacts to be relieved without changing the overall binding.[133] The inhibitors crystallised in the structures of RT presently available (see Figure 3.4 for the chemical structures of the inhibitors) are structurally diverse, and each has its own particular structural characteristics. However, a common binding mode suggestive of the shape of a butterfly has been proposed for the inhibitors of this pocket.[30,215] The DDM-based superimpositions provided an objective method to compare the binding modes of the inhibitors, illustrating the V-shape of the butterfly wings in Figure 3.49 (left), which are believed to play an integral role in π–stacking interactions with the hydrophobic residues of the pocket.[30,31]

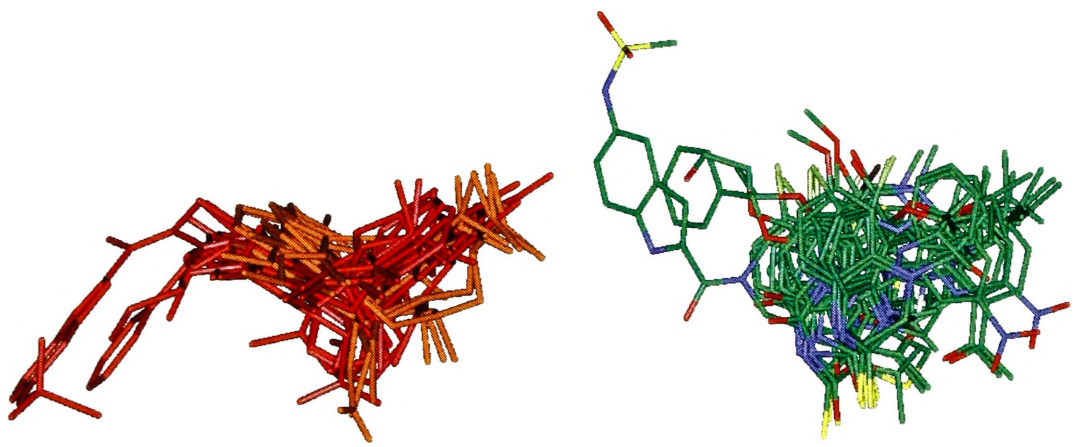


Figure 3.49 The crystallised non-nucleoside inhibitors of RT in the NNIBP by DDM-based superimpositions. (Left) Inhibitors coloured by the research groups that resolved the structures, CABM structures (orange), and OCMS structures (red). (Right) Inhibitors coloured by atom, oxygen (red), carbon (green), nitrogen (blue), sulfur (yellow), chlorine (pale green) bromine (brown).

A large proportion of the literature pertaining to RT has been based on intense studies of the nature and structural movements of the residues of the NNIBP, analysing how these could be better understood to refine and optimise current inhibitors,[138,216] and rationalise mutational resistance.[136,215,217] Our DDM-based studies were proposed not to follow in this same path, but rather to individually analyse a few residues with significant displacements from the conformation of the 1dlo structure, and to begin to examine the movements and extent of the pocket. This investigation would provide the basis for optimisation and *de novo* inhibitor design that could fully exploit all the potential protein interactions available and not be limited to those bonding interactions currently realised by known inhibitors within the NNIBP (Chapter 4).

3.7.2 Synopsis of the displacement of the NNIBP

The NNIBP lies seemingly between the palm and thumb subdomain, composed of residues from both enzyme subunits, and inclusive of several residues of the primer grip. Its position at the base of the thumb has been the basis for some speculation that it may function as a hinge between the palm and the thumb subdomains.[134] The binding of an NNRTI could distort the conformation of the primer grip and restrict the mobility of the thumb subdomain, believed to be necessary for interaction with the nucleic acid during recognition of the template-primer. In Figure 3.50, the location of the residues of the NNIBP is illustrated relative to the subdomains of RT and the PAS upon which it has such a profound effect. The residues of the NNIBP are listed in Appendix 7.6, and displacement data is tabulated in Appendix 7.13.

The binding of nucleic acid substrates appeared to do little in changing the conformation of the NNIBP as evidenced by the average displacements of 1rtd and 2hmi in Table 3.8. The average displacement of the catalytic 1rtd structure was actually less than that of the alternate 1hmv unliganded structure in both backbone and sidechain components by 0.14 Å and 0.25 Å respectively. The difference was even greater with the binary 2hmi structure, which showed the greatest displacement from the unliganded and substrate-bound structures with an average difference in distance to 1dlo of 1.48 Å. The black circle and dark grey diamond line traces of Figures 3.51 and 3.52 showed no remarkable differences from the unliganded structure, with occasional increases in displacement attributed to slight conformational changes in sidechains.

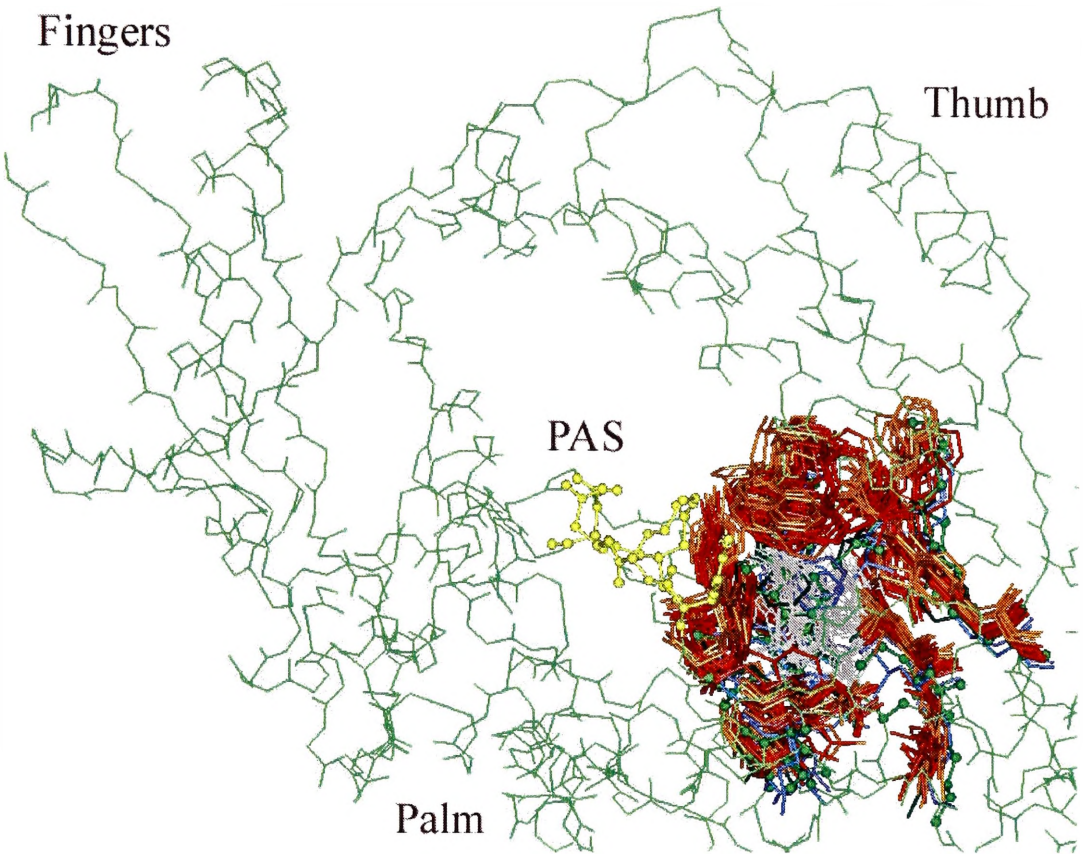


Figure 3.50 The displacement of the NNIBP, in the palm subdomain, upon substrate and inhibitor binding relative to the fingers subdomain, on the left, and the thumb subdomain (in the 'closed' conformation), on the right. The backbone of 1dlo is shown as a green line trace, the residues of the NNIBP is shown as green ball-and-stick, and the PAS is shown as the yellow ball-and-stick structure. Red and orange residues represent inhibitor-bound structures, light and dark blue structures represent DNA-bound structures. The inhibitors of inhibitor-bound crystal structures shown as grey lines.

average displacement (Å)		DNA-bound structures		inhibitor-bound structures		
	1hmv	ternary	binary	total	CABM	OCMS
entire residue	1.17	0.92	1.48	3.01	3.51	2.77
backbone	0.89	0.75	1.27	2.17	2.81	1.88
sidechain	1.39	1.04	1.64	3.65	4.06	3.45

Table 3.8 Summary of the average displacement between the 1dlo unliganded structure and the alternate unliganded (1hmv), substrate-bound (1rtd-ternary and 2hmi-binary), and inhibitor-bound structures for the entire NNIBP, and only the backbone or sidechain component. The three representative average displacements were calculated for all 19 inhibitor bound structures (total), and based on the research groups that derived the crystal structures; six from the CABM research group, and 13 from OCMS research group.

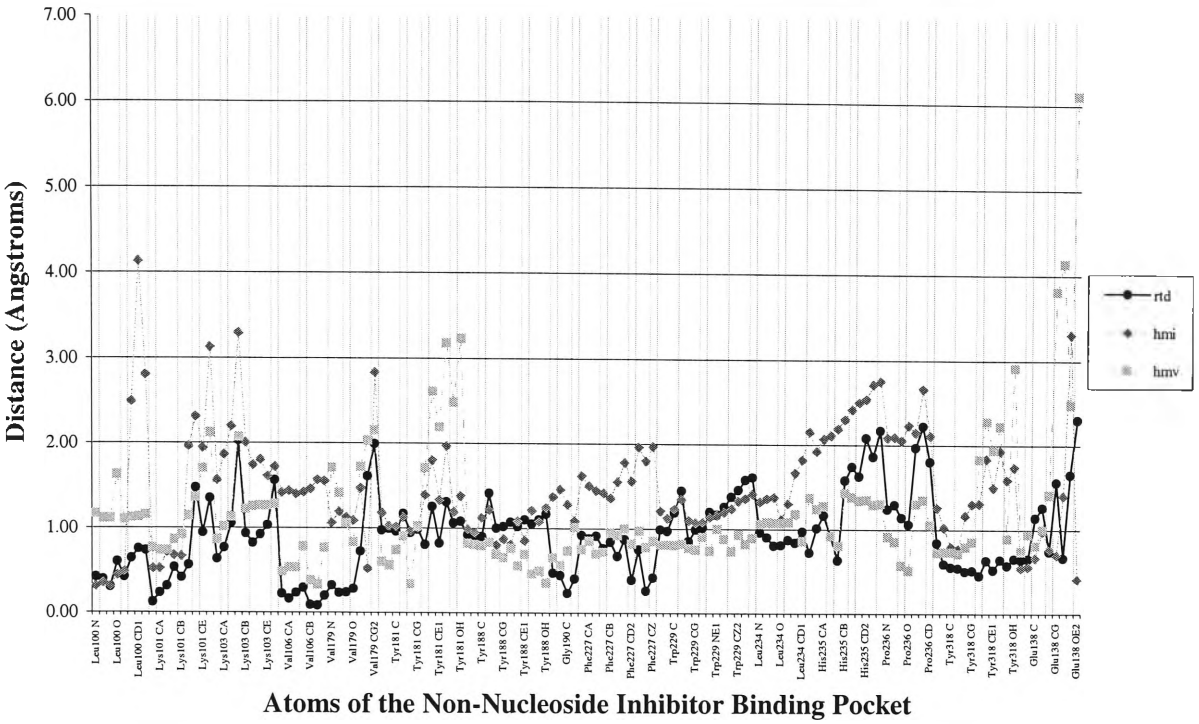


Figure 3.51 Graphical representation of the displacement between the unliganded (1dlo) structure, and the second unliganded (1hmv) and DNA-bound structures (1rtd and 2hmi) for the NNIBP.

Fundamentally, the binding of NNRTIs caused an approximate 30 ° differential twisting of the β12-β13-β14 sheet that resulted in the formation and expansion of the pocket.[30] However the NNIBP is composed of more residues than solely those of the β12-β13-β14 sheet, and their displacements from the conformation in the unliganded structure are most succinctly summarised in Figure 3.52. As the multi-coloured line traces clearly showed, the values of displacement in Table 3.8 consisted predominantly of the displacements of the sidechains of Tyr181 and 188, residues Phe227 and Trp229 of the primer grip, and to a smaller extent, Pro236 and Glu138. Removing the dominating aromatic residues from consideration brought the average displacement of the NNIBP in the inhibitor-bound structures to 2.20 Å, a much more modest displacement, but still distinct from the unliganded and DNA-bound structures. The four residues of 181, 188, 227, and 229 as a collective had an average displacement of 4.51 Å in the inhibitor-bound structures.

The NNIBP of the crystal structures of the CABM research group showed a greater displacement than that seen by the OCMS group by an average of 0.74 Å. Interestingly, the larger part of this difference in displacement was in the backbone component, 0.93 Å, rather than the sidechain component, 0.61 Å. The difference between the binding pockets of the two research groups was most likely to be a continuance of the greater displacements seen in the previous studied motifs, rather than due to the larger collective volume occupied by the red-coloured inhibitors of Figure 3.49 (left).

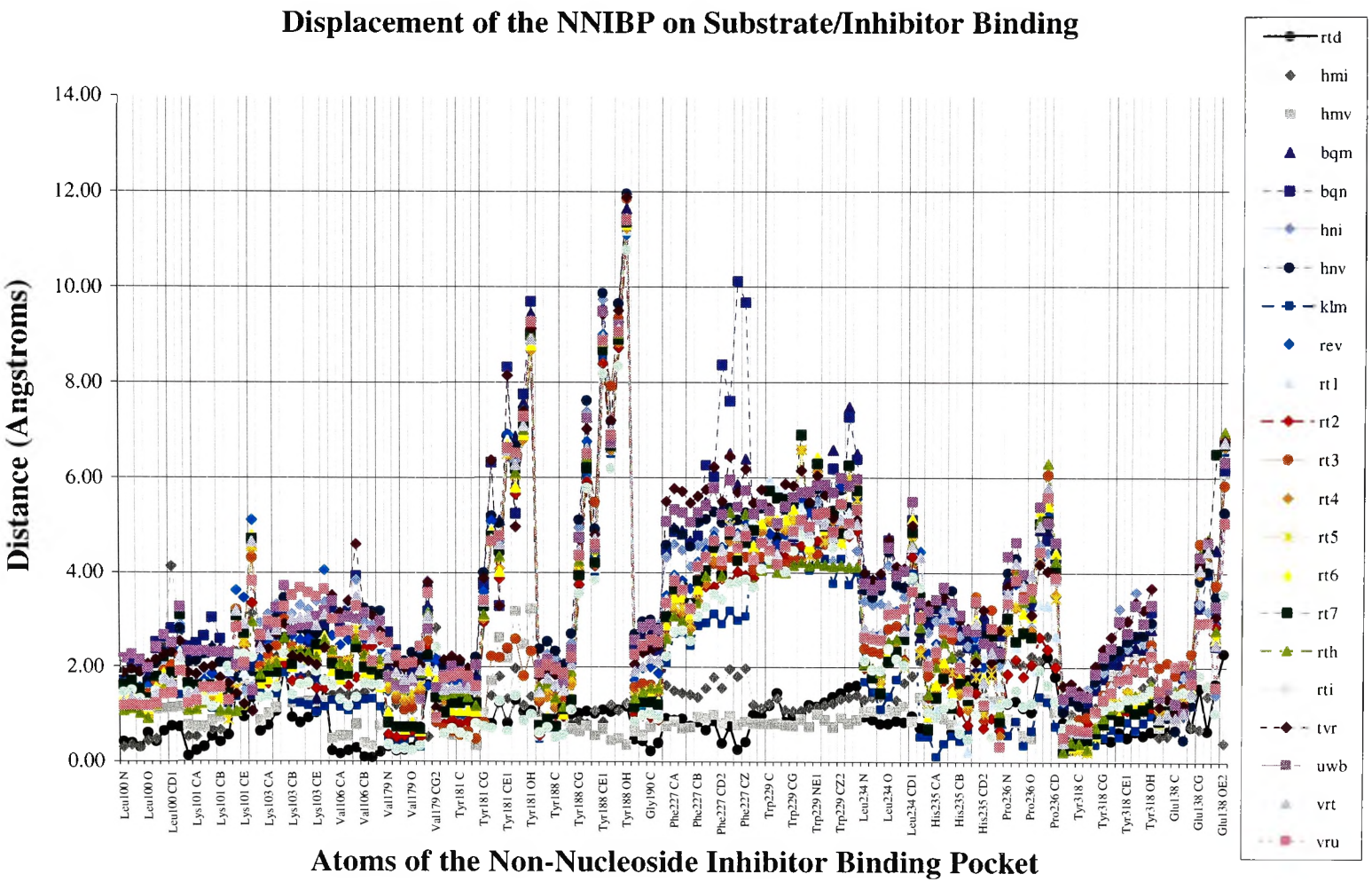


Figure 3.52 Graphical representation of the displacement between the unliganded (1dlo) structure and all liganded RT structures for the NNIBP.

3.7.3 Conformational changes for tyrosine 181 and 188

In the unliganded and DNA-bound structures, the sidechains of Tyr181 and Tyr188 of p66 pointed into the hydrophobic core and, as a consequence, the NNIBP did not exist. The binding of non-nucleoside inhibitors caused a rotation in the sidechains of both residues away from their positions in the hydrophobic core in the unliganded structure, to point upward toward the surface of the palm, and the PAS.

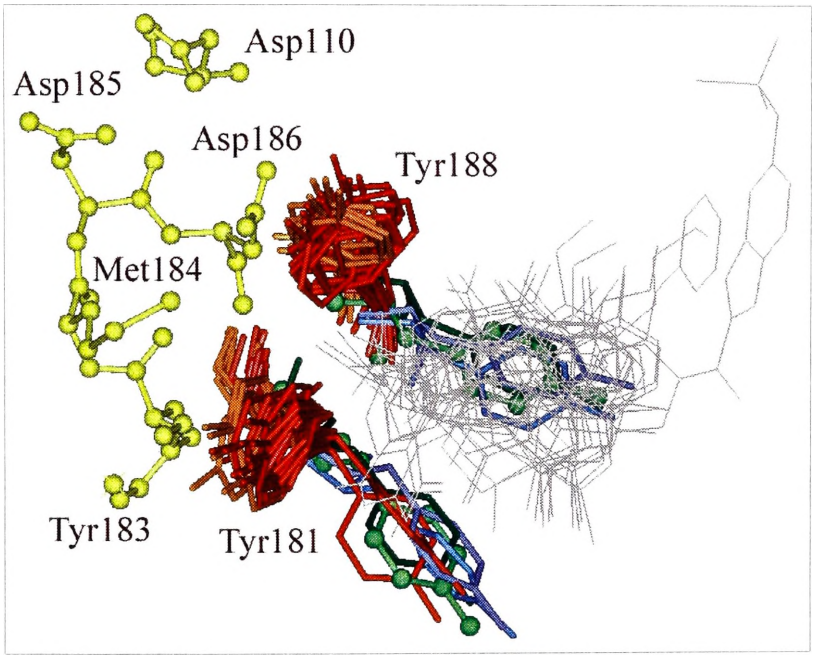


Figure 3.53 The displacement of Tyr181 and Tyr188 upon substrate and inhibitor binding seen relative to the PAS of 1dlo in yellow ball-and stick. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The inhibitors of inhibitor-bound crystal structures shown as grey lines. The fingers subdomain is to the left of the PAS and the thumb subdomain is to the right of the inhibitors.

At the bottom of Figure 3.53 the blue and green structures of the unliganded and DNA-bound structures can be seen to slightly occupy some of the space where a few of the NNRTIs bind in the NNIBP. A swivelling about the C_β angle in all the inhibitor bound structures, excluding 1rti and 1rt3, brought the sidechain of Tyr181 within approximately 1.5 Å of Tyr183 of the PAS. The lack of this swivel by the Tyr181 sidechain in 1rti and 1rt3 resulted in small displacements from the unliganded 1dlo structures as evidenced by the dark orange and pale green circle line traces of Figure 3.52 that lie close to the black and grey line traces. The sidechain torsion angle, χ_1, of Tyr181 in all the crystal structures was determined in Table 3.9. The average χ_1 of the unliganded and DNA-bound structures was –87.9 °, compared to 169.6 ° for all inhibitor-bound structures, excluding 1rt3 and 1rti, a difference of approximately 102.5 °. The average χ_1 value for 1rti and 1rt3 was –80.3 °, practically the same as the unliganded and DNA-bound structures.

PDB file	Tyr181	Tyr188
1dlo	-98.20	169.03
1hmv	-68.63	165.09
1rtd	-98.98	161.08
2hmi	-85.94	154.54
1bqm	169.14	-74.27
1bqn	159.95	-
1hni	172.25	-69.20
1hnv	-176.96	-77.10
1klm	169.68	-65.27
1rev	170.68	-72.63
1rt1	173.44	-59.66
1rt2	171.97	-63.53
1rt3	-80.28	-53.03
1rt4	166.64	-61.72
1rt5	167.28	-62.93
1rt6	167.53	-66.23
1rt7	175.86	-63.46
1rth	158.94	-69.49
1rti	-80.09	-65.07
1tvr	168.39	-65.09
1uwb	-	-81.63
1vrt	173.79	-69.11
1vru	171.22	-62.07

Table 3.9 Sidechain torsion angles of Tyr181 and Ty188 (N, C_α, C_β, C_γ).

No reasonable thorough explanation can be proposed for why the Tyr181 sidechain in these two crystal structures do not swivel. The structure of HEPT in 1rti was analogous to those of MKC-442 and TNK-651, but the χ_1 angles of 1rt1 and 1rt2, the respective crystals structures with these inhibitors bound, were on average 172.7 °, approximately 107.2 ° greater than in 1rti. The average difference between the χ_1 angles of the unliganded and DNA-bounds structures compared to most of the inhibitor-bound structures was 102.5 °. In Figure 3.55 (left), the comparison of the dark HEPT and the pale MKC-442 and TNK-651 inhibitors showed no altered binding modes or steric features of the inhibitors that could explain the difference in the sidechain conformation. The inhibitor of the 1rt3 structure was the nevirapine analogue, 1051U91. While the 1rth and 1vrt structures, that have 1051U91 and nevirapine bound respectively, had χ_1 values generally comparable with most of the inhibitor-bound structures, these were native enzymes, unlike 1rt3 which was an AZT-resistant RTMC mutant.[218] The mutations, two of which included Thr215 and Lys219, were known to give rise to conformational changes that propagate to the active site aspartates, as well as the NNIBP.[218] In Figure 3.55 (right), the 1051U91 and nevirapine structures all overlaid in a fairly common binding mode that provided little explanation for the position of Tyr181 in 1rt3. The phenyl ring in the 1rth and 1vrt structures was close enough to the left-hand aromatic ring of the NNRTIs in Figure 3.55 (right) to have staggered planar π-bonding interactions. The tyrosine sidechain in 1rti was an equal distance away from the same ring system of the 1051U91 inhibitor, but not orientated in a manner to allow any hydrophobic ring stacking interactions. This difference in the sidechain of Tyr181, among others observed in 1rt3, has been implicated in the discrimination between the inhibitor, AZT-triphosphate, and the natural dTTP substrate by the mutant enzyme.[218]

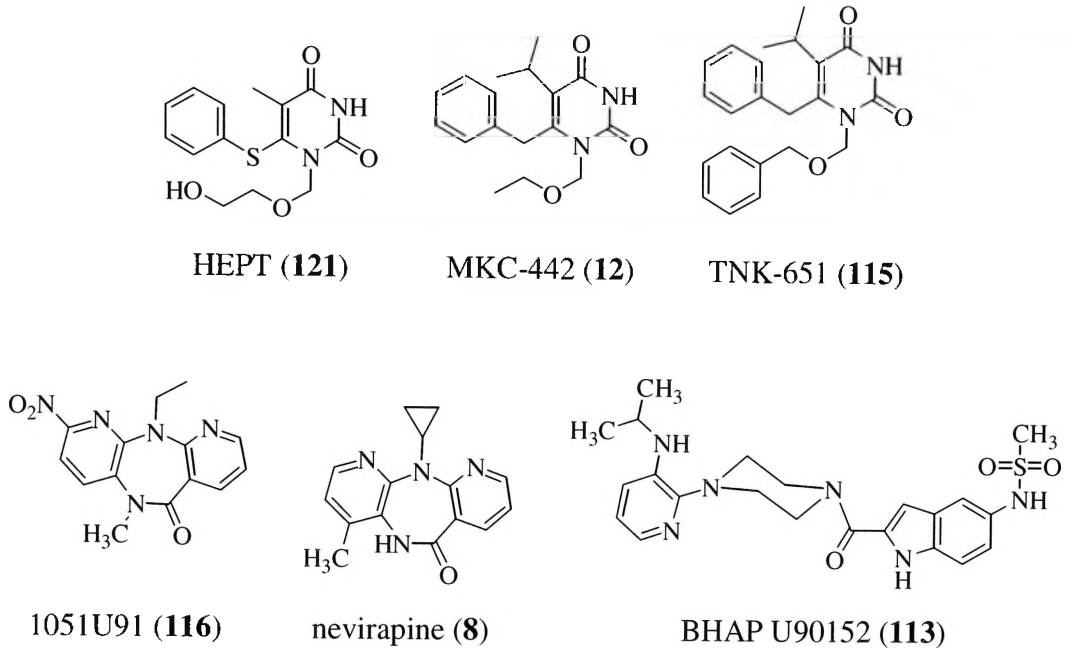


Figure 3.54 Structures of HEPT, MKC-442, TNK-651, 1051U91, nevirapine and BHAP U90152.

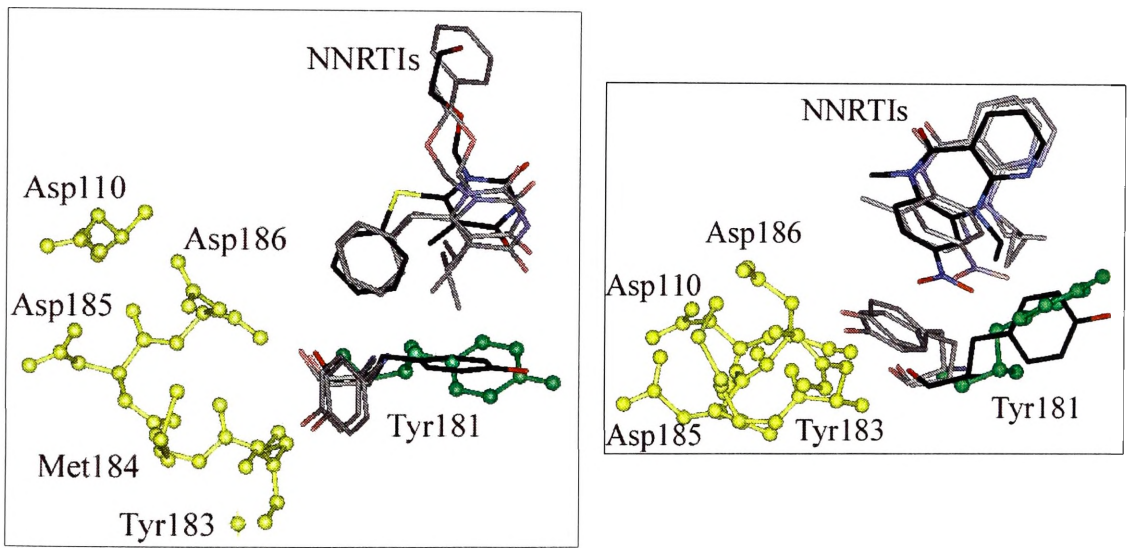


Figure 3.55 (Left) The conformational differences in the Tyr181 residue between the unliganded structure 1dlo (green ball-and-stick), 1rti (dark sticks), and 1rt1 and 1rt2 (pale sticks) seen relative to the PAS in 1dlo (yellow ball-and- stick). (Right) The conformational differences in the Tyr181 residue between the unliganded structure 1dlo (green ball-and-stick), 1rt3 (dark sticks), and 1rth and 1vrt (pale sticks) seen relative to the PAS in 1dlo (yellow ball-and-stick). Inhibitor-bound structures coloured by atom; oxygen (red), carbon (black), nitrogen (blue), sulfur (yellow).

Unlike Tyr181, all the inhibitor-bound structures of Tyr188 showed a change in the χ_1 angle to swivel away from the hydrophobic core of the pocket and upward toward the surface of the palm. The sidechains in the green and blue unliganded and DNA-bound structures also occupied some of the volume of the inhibitors in the pocket. The average change in χ_1 from the unliganded and DNA-bound structures to the inhibitor-

bound structures was 130.8 °, and allowed the aromatic sidechain to be positioned for possible similar staggered planar π-bonding interactions as in Tyr181, as well as bring the residue within approximately 1.5 Å of Asp186 of the PAS. These changes in the χ_1 of Tyr181 and 188 verified the previous observation that the sidechain torsion angles differed by about 120° in the structures derived in the presence and absence of inhibitors.[30]

3.7.4 Conformational changes for phenylalanine 227 and tryptophan 229

Phe227 and Trp229 form part of the primer grip and their distinct conformational changes have already been analysed to establish the changes induced by substrate and inhibitor binding (Sections 3.4.3 and 3.4.4). With respect to how they form the NNIBP, Figure 3.56 illustrates the shifting of the red and orange-coloured structures outward from the grey clustered inhibitors.

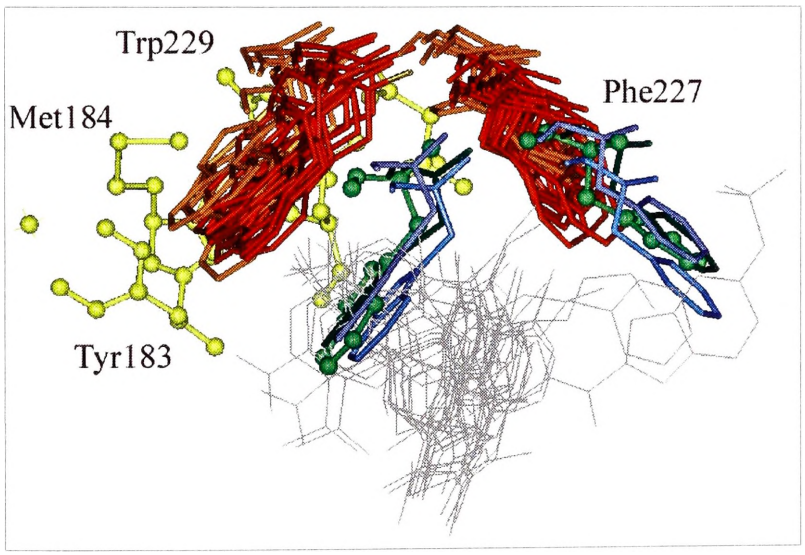


Figure 3.56 The displacement of Phe227 and Trp229 upon substrate and inhibitor binding seen relative to the PAS of 1dlo in yellow ball-and stick. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The inhibitors of inhibitor-bound crystal structures shown as grey lines. The fingers subdomain is to the left of the PAS and the thumb subdomain is to the right of the inhibitors.

The green and blue unliganded and substrate bound structures of Trp229 occupied a significant volume of the pocket where the grey line inhibitors were bound in the NNIBP of inhibitor-bound structures. Phe227 in the unliganded and DNA-bound structures did not superimpose over any of the volume occupied by the grey NNRTIs. The shifting of the red and orange-coloured inhibitor-bound structures, in this residue, served primarily to increase the size of the NNIBP, possible for entry into the pocket, rather than solely to accommodate the bulk of the inhibitor. Phe227 and Trp229, are conserved residues of

the pocket[219] and are believed to participate in significant π-bonding interactions with the NNRTI to help determine the specificity of binding.

3.7.5 Conformational changes for proline236 and glutamine 138

The conformational displacements of the Pro236 and Glu138 seen in the multi-coloured line traces of Figure 3.52 showed that, although they were not of the same consistency and magnitude as Tyr181, Tyr188, Phe227, and Trp229, they were markedly different from most of the remaining residues of the NNIBP. Pro236 was of particular interest as it is positioned at the mouth of the pocket and is next to His235 of the primer grip. This residue in the 1rt1, 1rt2, 1rti structures and most particularly in the 1klm structure showed little displacement from the conformation and position in the unliganded and DNA-bound structures. The grey inhibitor seen toward the top of Figure 3.57 was that of the BHAP U90152 (**113**, Figure 3.54) inhibitor of the 1klm structure. The similar conformation of Pro236 in 1klm to that in 1dlo was necessary to accommodate the methyl sulfonamide sidechain, as well as providing hydrophobic interactions to the inhibitor indole ring.[217] The proline sidechains of most of the inhibitor-bound structures were actually displaced forward at the mouth of the pocket from the unliganded structure, and occupied the volume of the indole ring of BHAP U90152 in the crystallised 1klm structure. This movement may constitute a narrowing of the mouth of the pocket and could affect how the association and disassociation kinetic rates of the inhibitor by

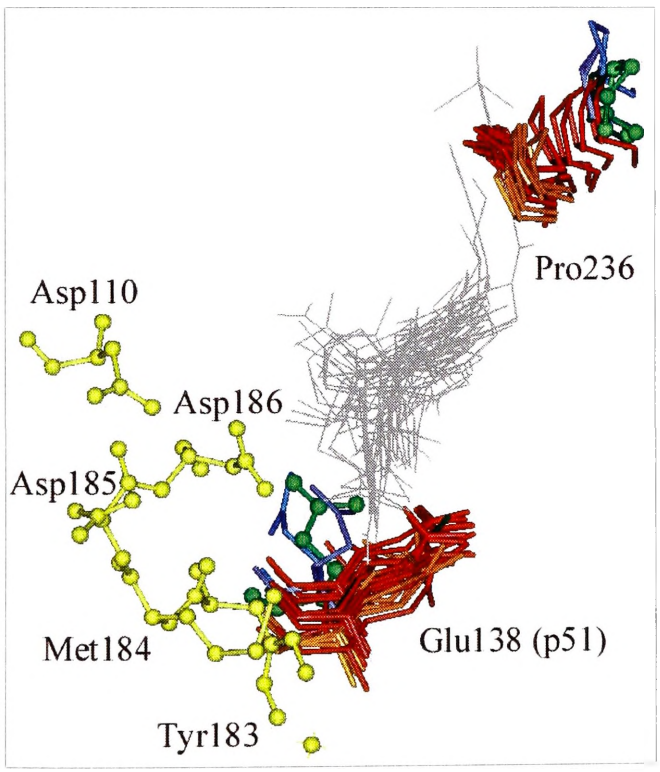


Figure 3.57 The displacement of Pro236 and Glu138 (p51) upon substrate and inhibitor binding seen relative to the PAS of 1dlo in yellow ball-and stick. 1dlo ball-and-stick (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red). The inhibitors of inhibitor-bound crystal structures shown as grey lines. The fingers subdomain is to the left of the PAS and the thumb subdomain is to the right of the inhibitors.

restricting its exit from the pocket. A comparison of the kinetic rates of the inhibitors to the space or volume of the mouth of the pocket would be worthy of note, but may be quite difficult as many more variables such as the bonding interactions within the pocket and the total energy of the inhibitor and the protein need to be considered. More detailed study of this residue could provide further understanding of the process of inhibitor entry into the pocket, and provide another bonding interaction for inhibitor optimisation that may not have been fully exploited thus far.

Glu138 is only residue of the p51 subunit to form part of the NNIBP. It does not lie near the collective volume occupied by the grey inhibitors of Figure 3.57, but its position and conformation could influence such residues as Tyr181 that can impact directly on the binding of the inhibitors. The dark green structure of 1hmv in Figure 3.57 and the light grey square line trace in Figure 3.52 revealed that the alternate unliganded structure had a much more similar conformation and position to the inhibitor-bound structures than to 1dlo and the DNA-bound structures.

3.7.6 The binding of non-nucleoside inhibitors

The stimulus for the DDM-based superimposition study of the NNIBP was predominantly as a basis for structure-based drug design studies, and the observations presented in this section of research represent only the initial stages of the research potential of this significant structural motif of RT. However, the results provide an exciting glimpse at the latent wealth of information for optimisation and *de novo* drug design, as well as understanding the mechanisms of resistance within the NNIBP.

Resistance mutations are thought to function primarily by altering the binding of NNRTIs to RT; that is, by the loss of favorable contacts between enzyme and inhibitor, and slower inhibitor binding due to changes at the entrance of the NNIBP. NNRTI resistance is believed to arise from a combination of loss of stabilising interaction, and the emergence of steric and thermodynamic barriers for drug binding, depending on the particular amino acid substitution involved.[220,221] The objective and rational approach of the DDM-based superimpositions provides an ideal method to study the possible conformations of the residues of the NNIBP, and so indirectly evaluate the flexibility of the pocket. Residues furthest away from the central volume occupied by the inhibitors could provide a measure of how far the pocket is able to expand to accommodate NNRTIs, and help in designing drugs that fit better and bind optimally to the enzyme.

3.8 Overall displacement trends and activity correlation

The DDM-based superimposition studies of the structural motifs of RT implicated in the inhibition of polymerisation, or RNase H hydrolysis, have detailed considerable and consistent displacement trends in both substrate and inhibitor-bound structures. Also revealed has been an insight into the complexity of the mechanism of activation and, in particular, inhibition of the enzyme. The multiple functions necessary for successful reverse transcription and the dynamic nature of these functions was likely to require the synchronous convergence of numerous structural elements of the enzyme, metal cations, template-primers and dNTPs. No single motif could be responsible for the entire function of the enzyme, nor for its inhibition by NNRTIs. The exact relationships between the active sites and the nucleic acid grips was likely to be important for communication throughout the enzyme as it reverse transcribes an RNA strand into DNA, before hydrolysing the copied template. Consequently, the binding of NNRTIs that induced the conformational changes detailed in this research was anticipated to alter these vital relationships.

The analogy of the opening of the hand has been previously used in literature to describe the structural changes of the primer and template grips upon inhibitor binding.[79,160,222-224] The results of the DDM-based superimpositions have supplemented this hypothesis to include the polymerase and RNase H active sites. The binding of NNRTIs caused the four structural motifs studied in Sections 3.3 - 3.6 to be displaced further away from the nucleic acid template-primer, essentially an opening or unclenching of the hand. The PAS was elevated up from the palm, away from the expansion of the elements of the NNIBP, to accommodate the non-nucleoside inhibitor. The primer grip was elevated up the plane of the thumb and, and even though it was displaced forward into the primer strand, the end result was a distancing away from its positioning in the unliganded and DNA-bound structures. The thumb subdomain was hyper-extended beyond the 'open' conformation of the DNA-bound structures, as were the fingers segments of the template grip, that were displaced deeper into the fingers domain and further out away from the nucleic acid. The RNase H active site also showed a gross displacement further away from both the primer-template complex, and the bulk of the p66 subunit.

3.8.1 Spatial relationships between structural motifs

The distance between the Gly231 C_α atom of the primer grip loop to the aspartates of the PAS in all the available crystal structures has already been measured and analysed (Section 3.4.6, Figure 3.28). A clear increase in this distance was seen in the inhibitor-bound structures as compared to the unliganded and DNA-bound enzymes, and this difference in distance must be noted as a possible factor in the consideration of the changes between motifs that may cause inhibition. The spatial relationship between the template grip motif, and primer grip and PAS were also considered in the inhibition of the polymerase activity. The segment of the template grip running across the fingers was analysed by measuring the distance between Lys73 C_α at the end of this segment, closest to the first finger, and Gly231 C_α of the primer grip loop (data tabulated in Appendix 7.13). The results were graphed in Figure 3.58 as a blue line diamond trace, with the different crystal structures coloured.

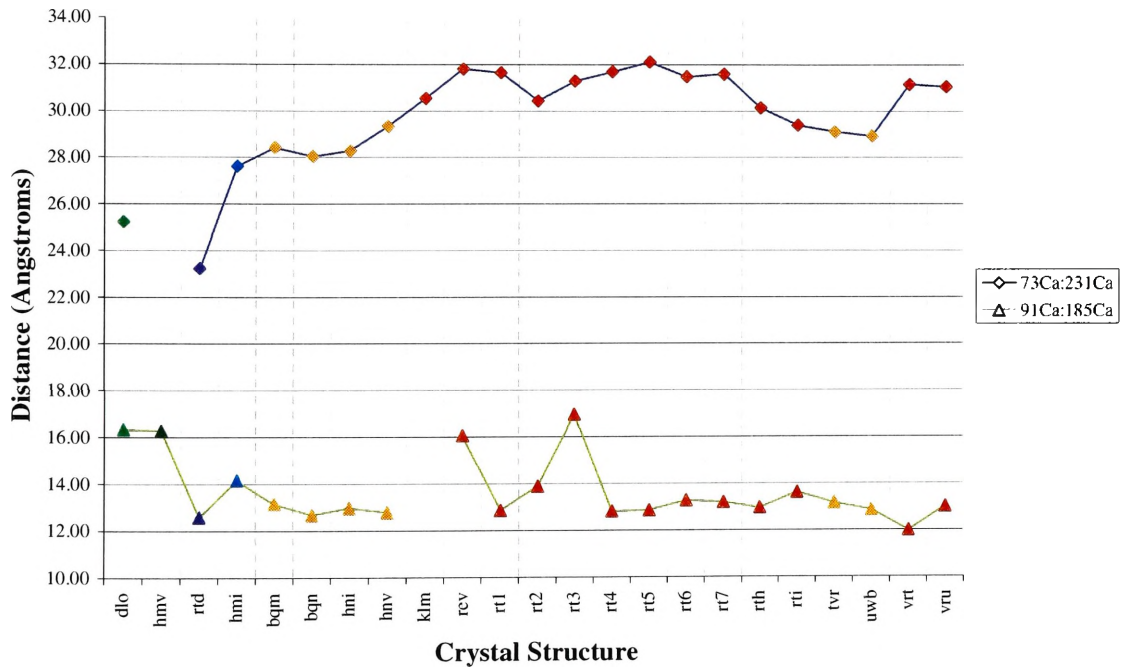


Figure 3.58 Graphical representation of the distance between fingers segment of the template grip to the primer grip (C_α Lys73 to C_α Gly231 – blue line diamond trace), and between the palm segment of the template grip to the PAS (C_α Gln91 to C_α Asp185 – green line triangle trace) in all of the crystal structures. 1dlo (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red).

The diamond line trace showed an increase in this distance in the red and orange coloured inhibitor-bound structures compared to the unliganded 1dlo (5.10 Å) and the catalytic 1rtd (7.11 Å). The relative similarity in the distance in the 2hmi structure was

due to a displacement of this segment of the template grip deeper into the fingers domain, and a downward displacement in the plane of the fingers.

The increase in the distance between these two motifs is likely to affect the communication necessary to correctly and efficiently base pair incoming nucleotides at the 3'OH terminus of the primer strand to the *n* nucleotide of the template strand that is presently being reverse transcribed. The greater distance between the two motifs primarily responsible for this positioning could mean that bonding interactions necessary for base recognition, if possible at all, would be extremely weak. Processivity as well as fidelity of replication are likely to be affected, as would the rate of chemical polymerisation and DNA replication

The function of the segment of template grip that runs across the palm subdomain has not been clearly determined, and as such, the significance of changes in spatial relationships between this motif and others of the enzyme can only be speculated at. The spatial relationship between Gln91 C_α of this segment of the template grip and Asp 185 C_α of the PAS was determined, and presented in the green line triangle trace of Figure 3.58 (data tabulated in Appendix 7.13). With the exception of the 1rev and 1rt3 structures, the inhibitor bound structures showed a decrease in the distance between this palm motif and the catalytic aspartate residue by an average of 3.30 Å compared to the two green unliganded structures. The two blue DNA-bound structures also showed an average decrease in this distance of 2.93 Å when compared to the unliganded structures, but were approximately equivalent to the distances observed for the inhibitor-bound structures, 13.37 Å compared to 13.00 Å respectively.

The similarity between the distances measured in the unliganded structures (16.32 Å and 16.27 Å) provided a strong basis with which to compare against the average distance in the inhibitor-bound structures, excluding 1rev and 1rt3, of 13.00 Å. The distances between the template grip palm motif and the PAS in the 1rtd and 2hmi structures, while reasonable similar, differed by 1.59 Å, providing a weaker point from which to compare the spatial relationships. This segment of the template grip is known to affect the activity of the PAS,[181] with long-range interactions that may require exacting specificity for proper communication. The decreases in distances seen in the DNA-bound structures may be necessary for conformational changes to a catalytically competent and active form. The similar decrease seen in the inhibitor-bound structures may appear to bring the enzyme closer to an active form, but there is also the

consideration that the spatial relationship may not solely be a matter of distance, but also direction. These rudimentary relationships studied provide little insight into the palm segment's role in polymerisation, other than noteworthy observations such as those of Figure 3.58. This palm motif could possibly provide a structural communication relay point between the PAS and the RNase H active sites, that can function in a concerted manner.[197] More studies into these spatial relationships may provide greater insight and yet may also further increase the complexity of the spatial equations necessary for enzyme activation and inhibition.

The relationships the RNase H active site has with the other motifs of RT are often neglected, with so much focus centred on the polymerisation reaction, with little regard to the complete function of the enzyme. The connection to the palm motif of the template grip has not been studied here, but both small and large relationships have been observed to the PAS and the primer grips. In the blue line circle trace of Figure 3.59, a small difference was noted between the inhibitor-bound structures and the unliganded and DNA bound structures (data tabulated in Appendix 7.13).

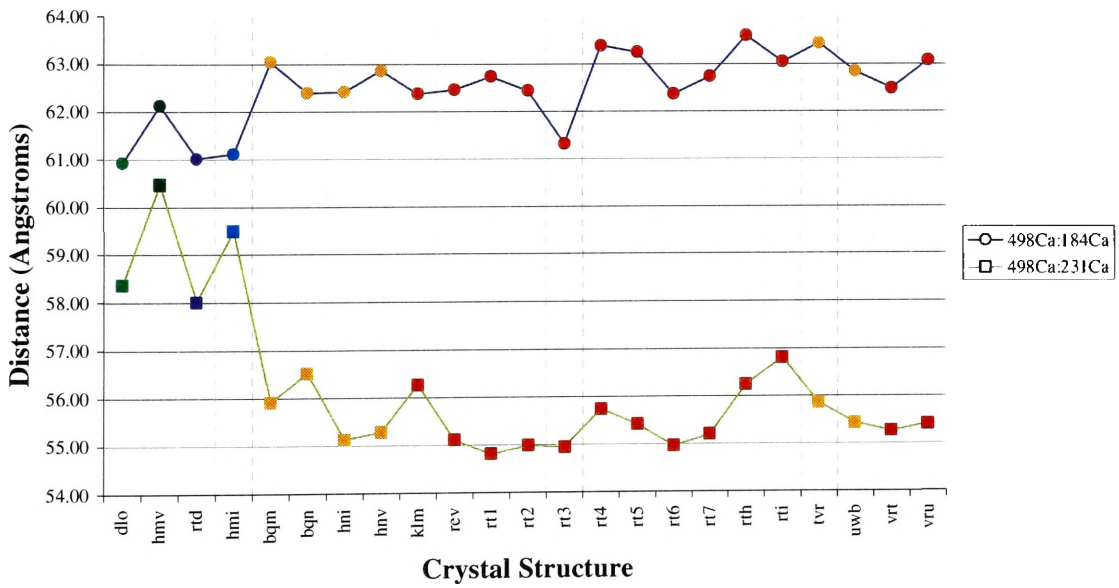


Figure 3.59 Graphical representation of the distance between RNase H active site to the PAS (C_α Asp498 to C_α Met184 – blue line circle trace), and between the RNase H active site to the primer grip (C_α Asp498 to C_α Gly231 – green line square trace) in all of the crystal structures. 1dlo (green), 1hmv (dark green), 1rtd (dark blue), 2hmi (light blue), CABM structures (orange), and OCMS structures (red).

The average distance between Asp498 C_α of the RNase H active site to Met184 C_α of the PAS in the inhibitor-bound structures was 62.74 Å compared to 60.93 Å in the 1dlo unliganded structure, and 62.12 Å in the alternate 1hmv structure. The mean distance in the substrate-bound structures was 61.06 Å. The average differences in the distances between the inhibitor-bound, unliganded and DNA-bound structures were all less than 2.0 Å, seemingly insignificant compared to some previously studied differences, and less than 4% of the distance measured between the two C_α atoms. However, the blue circle line trace of Figure 3.59 does appear to show an increased trend, supporting speculation that communication between the active sites could be in terms of precise spatial distances, with changes to those exacting distances affecting what could be communicated between the two sites. While the evidence presented, in reality, does little to corroborate this speculation, the challenge remains for further research to verify or refute these hypotheses.

The relationship of the RNase H to the primer grip motif was fortunately not as ambiguous, as seen in the green square line trace of Figure 3.59 (data tabulated in Appendix 7.13). The distance between Asp498 C_α and Gly231 C_α of the primer grip loop showed an average of 59.42 Å for the unliganded structures, and 58.75 Å for the DNA-bound structures. Interestingly, the unliganded structure of 1dlo showed greater similarity to the catalytic 1rtd structure (0.37 Å difference) than to the alternate unliganded 1hmv structure (2.10 Å difference), and greater similarity than between the two substrate-bound structures (1.48 Å). The average decrease in the distance between the RNase H and the primer grip in the inhibitor-bound structures compared to the unliganded structure was 2.83 Å, but quite definite as seen in Figure 3.59. The lack of equivalence between the unliganded structures made it difficult to more definitively establish this change in the spatial relationship, and correlate it to a mechanism of inhibition. Likewise, the difference between the DNA-bound structures made comparisons difficult, but this difference may be significant in another light, as a change in the spatial relationship to signify a conformational change from a nucleic acid bound inactive form to a catalytically active form.

An infinite series of measurements could be made to establish an infinite number of spatial relationships within the dynamic and multifunctional reverse transcriptase enzyme. The scope of such a project was beyond the field of this research, but there is no denying that such relationships do exist, and may very likely be the key to a

complete mechanism of activation or inhibition. This research has shown rudimentary evidence of associations and connections from the PAS to both nucleic acid grips, between the two active sites, the two nucleic acid grips, and between the RNase H and the primer grip. Further study will only serve to develop and augment these observed results.

3.8.2 Activity correlation to structural and conformational changes

One of the foremost hopes of this research was to find a correlation or mathematical relationship between the structural and conformational changes of RT, observed and quantified by the rational and objective DDM-based superimposition technique, and the inhibitory data of enzymes bound with NNRTIs. Previous attempts have been made to relate levels of polymerase activity to the measured distance between the primer grip and dNTP binding regions in modelled solvated wild-type and mutated RT enzymes, with and without a NNRTI complexed.[225] However, no other published study has included 19 inhibitor-bound crystal structures or correlated activity to inhibitor-induced conformational changes within and between structural motifs. The sometimes small changes within a structural motif, or between motifs, required that inhibitory data be precise so that variations in inhibitor structure could be differentiated. Unfortunately this endeavor was, with the presently available data and means, wishful thinking.

Several stumbling blocks make this objective presently infeasible. The multiple functions of the enzyme would require separate activity data for each of the functions studied, in particular RNA- dependent and DNA-dependent DNA polymerase activities, as well RNase H activities, in both native and inhibitor-bound enzymes. Other aspects of function need also be considered; processivity, fidelity, and phosphorolytic activity. Screens and assays would need to be consistent, with the same cell lines and experimental conditions used for all inhibitor testing. The impossibility of obtaining this data has already been previously observed,[24,26,27] and so such studies were not investigated beyond preliminary consideration with the crude activity data (Appendix 7.3 X-ray crystallographic structures of reverse transcriptase) acquired for the ligand and structure-based drug design studies of Chapter 4.

3.9 Integrated mechanism of activation and inhibition

No one study can hope to definitively and exhaustively analyse all the interactions of an enzyme to determine its mechanisms of activation and inhibition. There are numerous additional motifs and single amino acid residues of significance in RT that have yet to be studied, and may play consequential and critical roles in the enzyme function and dysfunction. This research has established the methodology by which any other structural motif can be objectively examined and its conformational displacement, upon substrate or inhibitor binding, quantitatively analysed. The active sites and motifs studied in this research have given an insight into the complexity of the reverse transcriptase enzyme, its multiple and dynamic functions, and the subtle conformational changes with which it carries out it biological functions and is inhibited by the binding of non-nucleoside inhibitors.

3.9.1 DNA *vs* RNA

The primary function of RT is to reverse transcribe double stranded proviral DNA from a single strand of viral RNA to ensure the genetic survival of the virus. This seemingly simple function requires an elaborate process of primer and template bindings, polymerisation, strand transfers, and RNase H hydrolysis (Section 3.1.2). However at each stage of the reverse transcription process, one basic yet crucial prerequisite is needed, the differential recognition of ribonucleic and deoxyribonucleic acids. A study of the differences between the structural features and forms of single and double strand RNA and DNA, RNA-DNA hybrids, and the primers of HIV-1, $tRNA^{Lys3}$ and the PPT, was beyond the scope and intent of this research, as was an analysis of the differences in nucleic acid conformation in RT (For reviews [162,163,168,174,175,211,226-229]). However, evidence that some functions of RT are sensitive and respond to the differences in RNA and DNA, provides a basis for the postulation that structural elements of RT are responsible for nucleic acid recognition, changing conformation to accommodate the correct form and function appropriately to that form.

Although reverse transcriptase uses both DNA and RNA templates in its natural replicative reaction, the enzyme has been shown to exhibit a marked preference for the correct RNA template.[5] RT preferentially binds to $tRNA^{Lys3}$ even in the presence of a 100-fold excess of other tRNAs,[230] and unerringly selects the PPT amongst all other

short RNA segments for synthesis of (+)-strand strong-stop DNA.[171,173,231] The nucleotide incorporation rate during DNA synthesis was found to be significantly greater on RNA than DNA templates, 10 to 15 nucleotides per second in contrast with 0.6 to 0.7.[178,232-235] Furthermore, the fidelity, or rate of faithful replication, during polymerisation was greater with RNA templates.[236,237] The rate of dissociation, as a counter measure of processivity, has been observed to be greater for duplex DNA than for RNA/DNA heteroduplexes.[156] The rate and specificity of RNase H cleavage has also been observed to differ depending on which substrate, DNA/RNA or RNA/RNA, was used.[127,195,203,238-240]

This wealth of evidence alludes to fine and subtle variances between the interactions of the enzyme with RNA and DNA that are augmented to produce distinctively different effects. Whether it is the structure of the nucleic acid itself that determines and induces the conformational changes of RT, or the structure of RT which identifies the nucleic acid and conformationally rearranges in response to govern the recognition process, or even a synergy of the two, can only be postulated. However, this discrimination and its suspected accompanying conformational changes are believed to be detrimentally and differentially affected by the binding of NNRTIs, influencing all the aspects of RT function. Consequently the answer to the question of how RT is activated and functions, as well as how it is inhibited, takes on an incalculable sum of variables and dimensions. When one considers that the RT enzyme, *in vivo*, is not believed to function in isolation, but rather aided by the viral nucleocapsid protein,[127] even more variables need to be considered in a mechanism of activation or inhibition. The nucleocapsid protein (NC) is believed to promote annealing between the $tRNA^{Lys3}$ primer and RNA template, stimulate the initiation of (-) strand DNA synthesis and strand transfer, and increase the efficiency of polymerase–independent RNase H activity, as well as the processivity of DNA synthesis.[3,127,195]

3.9.2 Snapshots and conformations

The results of the DDM-based superimposition have also helped form the tentative hypothesis that at the most simplistic level, RT and its elemental motifs exist in at least six basic global conformations.

1. E
2. E•RNA(template)•RNA(primer)
3. E*•RNA(template)•RNA(primer)
4. E•DNA(template)•RNA(primer)
5. E*•DNA(template)•RNA(primer)
6. E•I

(E – enzyme, E* - catalytically active enzyme, I – NNRTI)

Another four possible conformations may exist for the nucleic acid template-primers bound to enzyme-inhibitor complexes. The 1dlo and 1hmv structures are obviously representative of the unliganded E conformation. The 2hmi structure may be a close approximation of the E•DNA(template)•RNA(primer), with a homoduplex DNA in place of the DNA-RNA hybrid. It is likely to represent the enzyme bound to the nucleic acid before RT is activated and begins polymerisation, or the conformation of the enzyme between successive additions of dNTPs.[136,160] The 1rtd structure is believed to be a comparable snapshot of the E*•DNA(template)•RNA(primer) conformation, once more with a homoduplex DNA in place of the DNA-RNA hybrid.[136,160] All the inhibitor-bound crystal structures are believed to be representative of a diversity of E•I conformations. Naturally, there are expected to be differences due to the variations in the structures of nucleic acids (Section 3.9.1), but the basic hypothesis remains as such. The differences between the dsDNA in 1rtd and 2hmi, illustrated in Figure 3.60 (left), and the previously studied active sites of the two structures (Sections 3.3 and 3.6) support the postulation of differences between active and inactive conformations of RT. Most especially near the two active sites, the structure of the nucleic acids are seen to differ notably, and by association, it could presumed that the conformations of the motifs of the protein would also differ. They may adapt to each other, reciprocally making conformational changes to accommodate the other, for the mutual benefit of the virus.

With regard to the active site and nucleic acid grips studied in this research, this hypothesis of basic conformations of RT postulates that each motif has its own distinct and unique conformational features that distinguishes it from other conformations. Collectively, their individual 'snapshots', and the three-dimensional spatial relationships connecting them into one dynamic protein, form one of the six postulated global conformations.

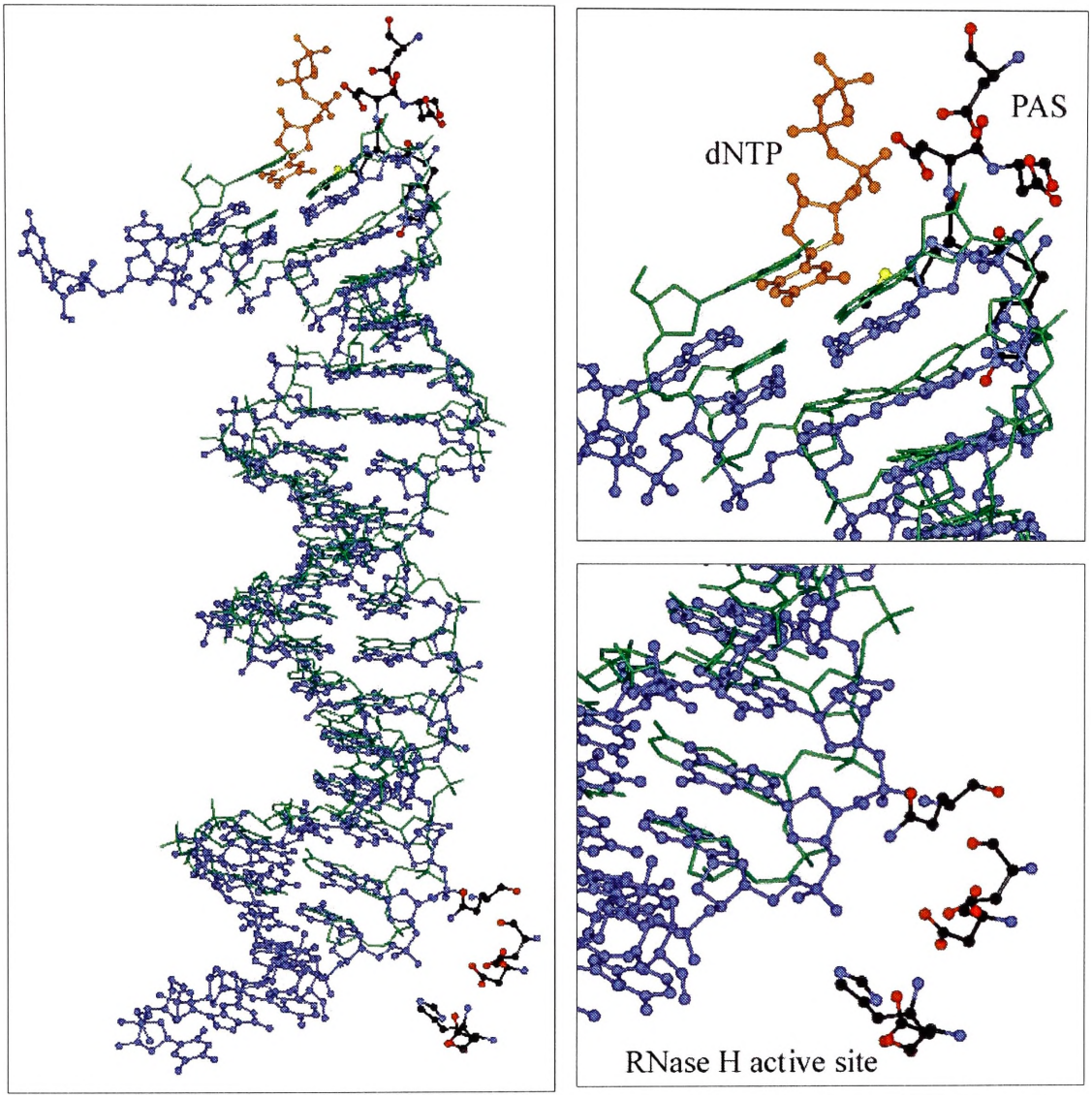


Figure 3.60 (Left) The structural differences between the dsDNA in 1rtd (blue ball-and-stick) and in 2hmi (green sticks), seen relative to the incoming nucleotide (orange) and the polymerase and RNase H active sites in 1rtd. Active site residues are coloured by atom, oxygen (red), carbon (black), nitrogen (blue), sulfur (yellow). (Right, top and bottom) Magnified views of the two active sites.

For ongoing studies into the mechanisms of activation and inhibition of RT, there is a need for more crystal structures or computer-generated models, in particular those with the natural RNA primers and templates bound, and structures with both nucleic acids and inhibitors concurrently bound. These structures will assuredly provide further insights into possible enzyme conformations and the ramifications of such structures. Such an inundating wealth of information may also provide enough information for distinct trends and patterns in conformational changes within motifs, and between motifs, to emerge. Despite the postulated complexity of the enzyme and its mechanisms, some simple structural and conformational relationships could be revealed, hopefully validating these present results and proposed conclusions.

Further study of the NNIBP using the DDM-based superimposition techniques should provide a means of characterising, and possibly quantifying, the conformational flexibility of the pocket, with the objective of using this information to predict inhibitor binding. The advantage to current drug optimisation and *de novo* inhibitor design would be an understanding of the mechanics of the pocket, possible movements and expansion that could allow the targeting of new binding interactions with conserved or newly realised residues of the pocket.

The information from the NNIBP could help in docking studies to predict drug binding modes and affinities. Current docking studies allow some flexibility in the binding site during minimisations, following initial docking, by allowing the sidechains to move and adopt new conformations around the inhibitor. However, the fundamental backbone of the binding site is restrained, producing a flawed simulation. The understanding and characterisation of the flexibility of the NNIBP will assist in the design of a more effective and efficient pseudo-pocket, representing the flexibility of entire individual residues. These residues could be allowed freedom of conformational change during docking minimisations, only limited by a spatial parameter defined by the least and most distant position the residues are found in recrystallised structures from a central inhibitor-binding position. Consequently, the potential of the pocket to bind inhibitors is more thoroughly and faithfully expressed.

Chapter 4

Integrated Drug Design

4.1 Computer aided ligand design

Computer aided ligand design (CALD) is a more accurate term for what is commonly known as computer aided drug design (CADD), and has two primary focuses; ligand-based and structure-based ligand design. The greater accuracy of the CALD term lies in the consideration that the research tools of the computational chemist are only able to stimulate and test ideas and concepts for the design of small molecules to bind specifically to biomolecules to induce a biological response.[78] Whether these small molecules or ligands progress through to be synthesised and undergo rigorous iterative optimisation cycles to pass tests for biological activity, specificity, bioavailability, lack of toxicity, medical need, manufacturing requirements and market potential to become a drug is another matter altogether. The objective of CALD is to use the diverse methodologies of computational chemistry to help discover novel or optimised chemical structures with properties that might qualify them to enter the drug discovery pipeline.

CALD secured its place in the drug discovery process predominantly by virtue of the economics of bringing a new chemical entity, or drug, from the laboratory bench to the patient, estimated to be anywhere from $200-$500 million dollars.[241] This cost includes not only the cost of the new drug, but also the cost of research on the multitude of failed lead compounds.[78] The tools of CALD are designed not only to conceive novel medicinal agents, but also to analyse their potential as drugs and determine problems with such issues as biological activity, specificity, bioavailability, and lack of toxicity as early in the drug discovery process as possible. Examples of the use of CALD are too numerous to detail in this short review, but include the design of inhibitors of HIV-1

protease[242] and reverse transcriptase,[243] beta-lactamase,[244] influenza virus sialidase,[245] and angiotensin converting enzyme[246] and have been well reviewed.[243,247-250]

CALD computational tools can be divided into two classifications; ligand-based and structure-based design tools. The terms receptor, binding site, active site, protein and enzyme are used interchangeably in this study, as are the terms ligand, and inhibitor, and molecule.

4.1.1 Ligand-based drug design

Traditionally ligand-based drug design, also known as analog-based drug design, has been used in the absence of three-dimensional structural information about the protein or receptor. It relies predominately on information about a series of compounds or inhibitors and their biological interaction with the protein. The value of the data depends on the specificity of the biological target. A series of compounds with a range of activities from an enzyme assay is much more valuable than compounds tested on whole cell assays. Ligand-based drug design utilises such computational tools as QSAR, CoMFA and pharmacophores to improve the biological activity of current drugs, and conceive new medicinal agents.

4.1.1.1 Quantitative structure activity relationships

Classical QSAR correlates the biological activities of drugs with their physicochemical properties or indicator variables that encode certain structural features such as lipophilicity, polarisability, electronic properties, steric properties.[251] QSAR can be used to gain a deeper understanding of the influence of lipophilicity, and ionisation and dissociation on the transport and distribution of drugs in biological systems.[252] The methods have been used to describe the affinity of ligands to their receptor in terms of different interaction forces between non-bonding molecules.[252] QSAR also allows the rational derivation and refinement of working hypotheses to aid drug design and lead optimisation.

The underlying assumption of all QSAR models is that the biological activity of different analogs is the sum of the contributions from individual substituents or groups of the parent compound.[252] Biological activities are defined as molar properties, on a logarithmic scale. All QSAR techniques assume that all the compounds being studied bind noncovalently to the same biological target; that structurally similar compounds

are similarly orientated at that common receptor or binding site; and that the dynamics of the system can be ignored.[253] QSAR methods differ in the way they describe the compounds, and in how they detect the relationships between the three-dimensional properties and biological activities of the compounds. However, all QSAR methodologies consist of three fundamental steps; molecular alignment, calculation of the molecular properties, and computational analyses of the relationships. An examination of the variations and technical details of these steps of QSAR has been reviewed by Oprea and Waller, and Greco *et al.*[253,254]

The term QSAR describes numerous different computational techniques, though the two most recognised are Hansch and Free-Wilson analyses. Hansch analysis, or the extrathermodynamic approach, of QSAR correlates the dependence of various biological activities on different physicochemical properties, while the Free-Wilson analysis of QSAR correlates it on the presence and absence of certain structural features.[252] In Hansch analysis, global properties such as partition coefficients, molar refractivity, lipophilicity, electronic, polar, and steric properties of substituents are correlated with biological activities.[252] In Free-Wilson analysis, values of 0 or 1 are assigned to indicator variables to code the presence or absence of certain structural features.[252] The potential and applicability of both methods can be augmented by QSAR models derived from a mixed approach of both Hansch and Free-Wilson parameters.[252]

The physicochemical properties used in QSAR are only limited by what can be described mathematically. Hydrophobic properties can include partition coefficients, solubility parameters, distribution coefficients, substituent constants, and hydrophobic fragmental constants.[255] Electronic descriptors can include Hammett constants, ionisation constants, ^{13}C and ^{1}H NMR shifts, electrostatic potentials, energies of highest occupied and lowest unoccupied molecular orbitals, and atomic net charges. Steric descriptors can include molar volume, molecular weight, van der Waals radii and volumes, and molar refractivity.[255]

Quantitative models or regression equations can be derived, and the significance of the complete or individual terms of the model can be tested by linear or nonlinear multiple regression analysis, discriminant analysis, principal component analysis (PCA), or partial least squares analysis (PLS).[252] The relevance and legitimacy of the QSAR

model can be tested by cross-validation, or lateral validation.[252] In the validation process the QSAR model is compared with closely related biological systems.

A classical QSAR regression equation is described in Scheme 4.1, where –log [Conc] is the negative logarithm of the concentration needed to evoke a biological activity. The coefficients, a, b, c, and so forth, are determined by regression analysis.[254]

Scheme 4.1

$$-\log[\text{Conc}] = a\left(\begin{array}{c}\text{steric}\\ \text{descriptor}\end{array}\right) + b\left(\begin{array}{c}\text{electronic}\\ \text{descriptor}\end{array}\right) + c\left(\begin{array}{c}\text{hydrophobic}\\ \text{descriptor}\end{array}\right) + \cdots\cdots$$

Three-dimensional QSAR methods are those that locate structural properties based on the three-dimensional structure of the modelled compounds.[254] These methods attempt to identify spatial regions for complementary molecular properties that match the binding site.

The quality of a QSAR analysis, its ability to predict values, such as the biological activity or toxicity, of compounds it has not seen, its capacity to explain the mechanism of action of individual ligands at the binding site, and its pharmacological relevance, define its value as a drug design tool.[254] The fulfillment of these criteria is as much dependent on the quality of the data or training set, as the QSAR method used, and this prerequisite is recurrent in all other drug design techniques.

4.1.1.2 Comparative molecular field analysis (CoMFA)

Comparative molecular field analysis (CoMFA) is a widely used tool for the study of three-dimensional quantitative structure-activity relationships, predominantly used in the field of ligand-protein interactions, to describe affinity and inhibition constants.[251] Originally developed by Cramer *et al.*,[256] CoMFA relates the biological activity of a series of molecules with their steric and electrostatic fields sampled at grid points defined by a three-dimensional box around the molecule.[257] It is an important precondition of these molecules that they all interact with the same receptor, enzyme, ion channel, or transporter in the same manner, with identical binding sites and in the same relative geometry.[251]

The descriptors, or field values, for CoMFA are usually calculated by means of different atomic probes such as a carbon atom, a positively or negatively charged atom, a

hydrogen bond donor or acceptor, or a lipophilic probe.[251] Classical CoMFA steric fields are calculated as Lennard-Jones potentials and the electrostatic fields are calculated as coulombic potentials.[257,258] Other CoMFA fields can include hydropathicity, hydrogen bonding, molecular orbitals, similarity indexes, and desolvation energy fields.[254] These field values are calculated at each grid point, and present the energy values that the particular probe would experience in the corresponding position of the three-dimensional lattice.

A relationship is developed between the independent variables, such as the steric and electrostatic potentials, and the dependent variable, such as biological activity or binding affinity, using a partial least squares regression method, that produces a regression equation with possibly thousands of coefficients.[251,257] This model equation, or equations, is most often graphically presented as a set of isocontour maps. These maps indicate the regions where the variation in steric and electrostatic properties of different molecules in a data set correlate with the variation in biological activity, or binding affinity.[251,257]

One of the biggest challenges in CoMFA studies is alignment, or superimposition, of the molecules of the test set, that is, the specification of which part of one molecule relates to a particular part of another molecule.[258] This can be complicated in cases of seemingly closely related analogs having different binding modes.[251] Alignment rules need not directly reflect the relative molecular orientations in the binding site of the receptor or enzyme. X-ray crystallographic, multidimensional NMR, and docking alignments of the ligand in ligand-protein complexes have been used for CoMFA.[259-261] However, they often fail to produce models that are as predictive as alignments based on shared substructure elements.[258] Superimposing common substructures tends to wash out field variations at the corresponding coordinates and focuses the analysis on those positions where structural variation is most relevant.[258,260] Thus CoMFA alignment can be evaluated by either a pharmacophore hypothesis of common substructures, a receptor site model from crystallographic, spectroscopy or docking data, or by a hybrid approach.[254]

Related techniques include comparative molecular similarity indices analysis (CoMSIA), and comparative molecular moment analysis (CoMMA). CoMSIA was devised to overcome rapidly changing steric fields near the van der Waals radius of an atom that result in ambiguities and fragmentation of the contours at these points.[254]

Contour maps from CoMSIA are smoother, and more coherent, and thus are easier to interpret.[251,262] CoMMA has the advantage that molecules do not need to be superimposed.[263]

4.1.1.3 Pharmacophores

A pharmacophore can be defined as the critical three-dimensional arrangement of points in space, each having some specific property such as charge, hydrophobicity, or hydrogen bonding capability, that is responsible for creating patterns of atoms and charge distributions that allow recognition by, and binding to, the target receptor site(s).[254,264] The most traditional method of pharmacophore characterisation begins with the identification of crucial functional groups in a ligand which have been observed to bind to a receptor or enzyme binding site.[264] Numerous conformations of the ligand may exist, and those in which the balance between the deformation energy to fit into the receptor and binding energy to the receptor is such that binding is thermodynamically permitted are examined in the pharmacophore analysis.[264]

A second approach to the generation of a pharmacophore involves the analysis of the receptor or active site. Information about the active site can be in the form of mutation data of single amino acid residues of the site that reveal their significance to receptor or enzymatic activity.[264] Increasingly, receptor site information comes in the form of x-ray crystallographic and NMR spectroscopic data of the protein structure.[264]

One of the most powerful applications of pharmacophores is in the mining or searching of electronic databases of three-dimensional structures to determine structures that contain the features of the pharmacophore that can elicit the desired biological response.[246] During such searches, the spatial dimensions of the pharmacophore are used to identify compounds in the database that contain the pharmacophoric features.[264] Multiple conformations of these compounds are generated and examined to determine if they fit the pharmacophore within reasonable energy terms. Pharmacophores can also be used for *de novo* generation of structures (Section 4.1.2.2).

Pharmacophores are invaluable tools for the identification of lead structures that can be examined for the appropriate biological activity, then chemically modified so as to improve the geometry, incorporate new binding interactions, or optimise their transport properties. Iterative modifications of this lead compound, guided by refined pharmacophores and molecular modelling, produce potential drug candidates with

optimal biological activity and minimal negative properties faster and more efficiently than traditional medicinal chemistry approaches to drug design.[264] However, despite the power of pharmacophores in rational drug design, their potential is often curbed by the structural diversity of the drugs that have spawned them. Any new drugs to emerge from these design principles can only make use of the same ligand-protein interactions as the parent drugs may have had.

4.1.2 Structure-based drug design

Advances in structure determination methods over the last couple of decades have supplied an increasing abundance of information on the three dimensional structures of proteins.[249] As of November 2000, the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank contained 13 672 structures; 11 232 derived from x-ray crystallographic techniques, 2138 from nuclear magnetic resonance studies, and 302 theoretical models.[265] Parallel with this growth was the development and application of new computational approaches to exploit this structural information in the drug discovery process.

Structure-based drug design is essentially the design of chemical structures to fit a macromolecular target of known architecture.[266] Knowledge of the three-dimensional steric and electrostatic environment of the active site of a receptor, and possibly the interactions between the residues of the active site and a bound ligand, can enable lead compounds to be modified while preserving the intermolecular interactions that confer binding affinity and specificity.[77] This also allows the design of totally novel ligands based upon shape and function complementarity with the known receptor structure. The two principal tools of structure-based drug design are molecular docking and *de novo* design; both are based on the analysis of protein structure, and ligand-receptor interactions.

4.1.2.1 Ligand docking

Molecular docking studies attempt to find the binding mode of a ligand in a ligand-receptor complex by searching the conformations and relative orientations of the ligand and receptor for a geometry with a favorable binding energy.[267] Knowledge of the bioactive conformation and binding mode of a ligand aids in the rational explanation of some of the features of structure-activity relationships for the ligand. Docking

programs can also screen databases of small molecules to identify ligands that will fit the target receptor site, to be used as leads for further drug development.[268]

Accurately modelling ligand-receptor interactions is difficult because of the large number of intermolecular interactions and degrees of freedom available to the ligand-protein complex. Three rotational and three torsional degrees of freedom are available between the ligand and the protein, as well as internal torsional degrees of freedom within the ligand and the protein.[269] Consequently, docking algorithms can be divided into four classes depending on the degree to which they consider molecular flexibility. Algorithms can regard both the ligand and protein as rigid molecules, the ligand as a flexible molecule while keeping the protein rigid, the ligand as a rigid molecule with the protein flexible, or with both ligand and protein as flexible molecules.[270-273] Conformational flexibility of the ligand is often incorporated into docking by calculating multiple conformations of the ligand, and then docking each individual conformer as a rigid ligand.[268,274] Conformational flexibility in the protein is often considered by holding the protein backbone rigid while allowing systematic variation of the amino acid side chains amongst a predetermined set of conformations determined from analysis of high resolution protein structures.[275] Ideally, the consideration of molecular flexibility should involve molecular mechanics, or molecular dynamics, however this is computationally infeasible for practically all ligand-receptor systems.[273]

Continuing developments in molecular docking will improve how the flexibility of the ligand and the protein are considered, as well as improving the methods for scoring the receptor binding affinity of each ligand relative to the others.[267] Molecular docking can be an effective tool for assessing the prospect of a ligand binding to a receptor, as well as for screening electronic databases for potential new ligands. However, this is provided that the intensive theoretical calculations estimating the binding affinity, and scoring functions are fast enough to be practically useful.[273]

4.1.2.2 *De novo* ligand design

While *de novo* ligand design can also be used with ligand-based pharmacophores, its prevailing use has been in structure-based studies.[247] *De novo* ligand design includes three basic computational methods, fragment positioning, molecular growth, and database search methods.[268] The fragment positioning method determines the energetically favorable binding site positions for various functional groups or chemical

fragments, and includes such programs as GRID®,[276] and LUDI®.[277,278] The optimal molecular fragments are then clustered and connected to form chemically conceivably and synthesisable candidate ligands.[268] The diversity of the resultant ligands is only limited by the variety of chemical moieties or fragments available.[279] Another advantage of this method is that the synthetic accessibility of the ligands can be controlled at the fragment connection stage.

In molecular growth methods, a seed atom or fragment is placed in the binding site and a ligand is successively formed by bonding another atom or fragment to it.[268] The ligand is scored as it is built, based on its chemical complementarity with nearby atoms in the binding site of the receptor active site. The disadvantage of this method is that the results are often very much dependent on the initial positioning of the seed atom in the binding site, and many of the resulting ligands may be too difficult to synthesise.[268]

The final computational method incorporates the use of database searching techniques to extract existing molecules, from an electronic database, that can be docked into the binding site with the desired fragments or functional groups in their optimal position. Programs such as Unity 3D®,[280] and the recent Structure-Based Focusing module of Cerius[2®],[281,282] can be used to determine or combinatorially generate possible structure-based pharmacophores or queries. These are a collection of pharmacophoric features placed at optimal positions to the residues of the three-dimensional active site, where a functional group from a ligand could be placed to form an intermolecular interaction with the residue of the active site. They can then be used to search databases in the same fashion as ligand-based pharmacophores.

Another important aspect of *de novo* design is the scoring function that allows some judgement to be made on the quality of the ligands proposed.[247] It is used to screen large numbers of structures or fragments to determine their fit in the receptor, and also used to rank the binding affinities of a few selected putative ligands before the determination is made on which is the best to synthesise and use as a lead compound.[279] Developments in *de novo* ligand design include the optimisation of these scoring methods to better assess the binding affinity of ligands.[249,279] Further advancements include the consideration of protein flexibility in ligand design, and the structural changes induced by the ligand onto the protein, and *vice versa*.[249,279]

Structure-based design strategies have the superior advantage that they are not limited or overly influenced by the structure of current drugs and inhibitors. All of the possible

ligand-protein interactions in a receptor site may be studied and utilised in the drug design process of novel medicinal agents, including new interactions that may avoid resistance mutations of the active site.[33] The potential of these structure-based design tools is limited by the need for nuclear magnetic resonance, x-ray crystallographic, or homology model data of the protein. However, the increasing challenge to understand and model the nature of protein-ligand interactions, and the need to accelerate the drug discovery process, is providing a driving force for the development of advanced methods for structure-based design able to consider these difficulties current being faced in drug design.

4.1.3 Integration of ligand- and structure-based drug design techniques

The complexity of any natural system is such that only one or a few facets of its complexity can ever be modelled at any one time. For too long, the understanding of protein ligand interactions, and consequently the drug design process, has been a case of two different facets of a natural system being modelled and studied separately; ligand-based and structure-based. This was because the entire puzzle has been too immense to study as a whole, and because parts of the puzzle have been missing. Although the immensity of the puzzle has not changed, more pieces of the puzzle have been revealed, as have more computational drug design techniques. This research was embarked upon to endeavour to bring together ligand and structure-based design techniques, to exploit the strengths from these two design methodologies; and compensate for their shortcomings and limitations.

Other combinations of computational tools have been utilised to augment their individual capabilities; 3D QSAR and receptor modelling,[283] pharmacophores and molecular docking,[284-286] pharmacophores and pseudoreceptor modelling,[287,288] and pharmacophores and 3D QSAR with excluded volumes from x-ray crystallographic structures.[289,290] In this undertaking, the central foundation is based on the concept of pharmacophores; classical ligand-based and emerging structure-based pharmacophores, and their fusion into one all-encompassing pharmacophore.

4.1.3.1 Strategy for the integration of ligand- and structure-based pharmacophores

Our proposed integration of ligand and structure-based drug design is illustrated in Figure 4.1. It is based on the fundamental concept of pharmacophores (see Section 4.1.1.3 Pharmacophores).

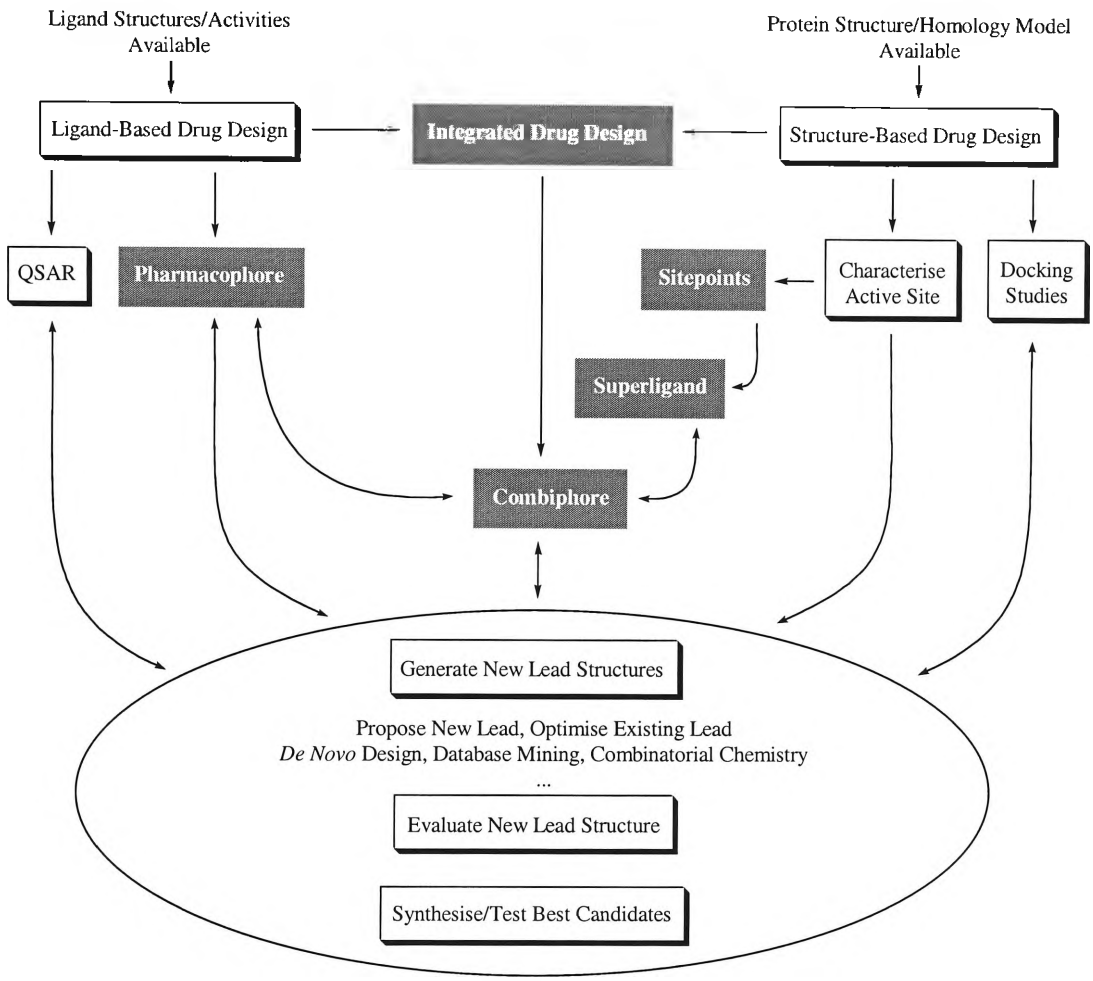


Figure 4.1 The integration of ligand-based and structure-based drug design strategies.

The strategy proposes to develop a traditional ligand-based pharmacophore with a standard tool of the ligand-based design, such as Catalyst®.[291] From structure-based design techniques, the receptor active site will be characterised by such programs as Unity 3D®,[280] LeapFrog®,[292] LUDI®,[293] or GRID[276]. These programs determine the spatial coordinates of 'site points', or points in space that describe the potential location and type of atoms or chemical moieties of a ligand, which would make favorable interactions with one or more of the protein atoms, as illustrated in Figure 4.2. Site points provide a measure of all of the potential ligand-protein interactions in a receptor; not just those previously identified by known ligands. As there can be hundreds of

sitepoints produced for a typical receptor site, they need to be clustered or averaged to locate the optimal position for a feature on a putative ligand molecule.

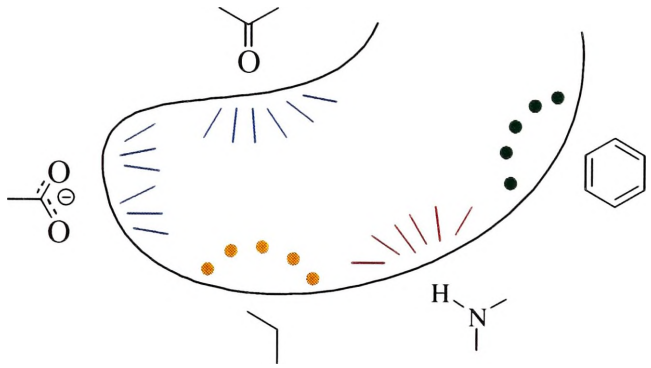


Figure 4.2 The generation of site points. Green dots represent where a hydrophobic aromatic chemical function may lie on a ligand, the red lines represent a hydrogen bond acceptor functionality, orange dots present hydrophobic aliphatic functionality, and the blue lines represent a hydrogen bond donor moiety. Typically, the hydrophobic aromatic and aliphatic site points are designated with one type of site point encompassing the two different interactions. Site points presenting positive charges, not shown, can also be generated.

Site point features will then be assimilated to form the superligand. Our proposed superligand is a structure-based pharmacophore that represents the largest and most optimal ligand that could possibly fit into the receptor or binding site. It will utilise all favourable and conceivable binding interactions with the protein, and will also include accurate excluded-volume information from the receptor structure. It does not take into account possible obstacles such as entry into a binding pocket. The structure of the receptor active site can be derived from x-ray crystallographic and NMR data, as well as from homology models.

The proposed combiphore will then be created by a merging of the classical pharmacophore derived from ligand-based design, with the superligand derived from structure-based methods. It would represent the most comprehensive synergy of ligand and structural information to date. The combiphore will, hopefully, employ all the strengths from these two design methodologies; and try to compensate for shortcomings and limitations in present design techniques.

In a supplementary attempt to try and compare the contrasting ligand- and structure-based techniques, only the ligands from x-ray crystallographic enzyme-inhibitor complexes will be used for the ligand-based pharmacophore, while the corresponding crystal structures will be used in the generation of the structure-based superligand. This will ensure that neither technique has a biasing effect on the final combiphore. And

although this may mean the number of the ligands in the ligand-based pharmacophore would be extremely small, the usual uncertainty of whether all the ligands bound at the receptor site or not would be removed.

Reverse transcriptase is an attractive drug target as it is an essential and unique enzyme in the replicative of HIV-1 (see Chapter 1). It is an excellent example to evaluate the potential power of our combined ligand- and structure-based pharmacophores for drug design as it is a well-documented protein. There are available numerous high-resolution crystal structures of RT with a variety of inhibitors bound in the non-nucleoside inhibitor binding pocket (NNIBP). These inhibitors, along with a multitudinous series of analogues, also provide biological activity data that has already been used extensively in ligand-based drug design as recently reviewed by Garg *et al.*[294] The availability of both ligand- and structure-based data will also allow for the comparison between the established pharmacophore methodology and the developing superligand concept, and hence validation of the superligand as a drug design tool. The inherent flexibility of RT and its propensity for mutation to evade inhibiting drugs also poses a challenge to rational drug design, which may be tackled by this new design approach.

4.2 Ligand-based pharmacophore generation with Catalyst

The first stage of the integrated drug design process involves the generation of a typical ligand-based pharmacophore, using the Catalyst® software program (Molecular Simulations Inc.).[291] Catalyst consists of three modules.[250] The first of these is the catConf module that computes an electronic representation of the structure of the ligands that emphasises coverage of conformational space. The second catHypo HypoGen module calculates chemical feature-based pharmacophore models from a training set of compounds with biological activity data, usually IC_{50} or K_i values. A related module called HipHop can, calculate pharmacophore models with common features in a training set without activity data. The third module is catDB, a compound database building program that can be searched with several kinds of queries, such as pharmacophore and compound shape models derived from catHypo and HipHop.

Catalyst pharmacophore models are called hypotheses, and the terms 'pharmacophore', and 'hypothesis' are used interchangeably in this study. Hypotheses are constructed from five types of ligand-receptor binding interactions; hydrophobic, hydrogen bond donating (HBD) and accepting (HBA), salt bridges, and aromatic ring π stacking

interactions.[250] These interactions are termed the features of a hypothesis, and are represented as points in space surrounded by a tolerance sphere. A hypothesis is a collection of these spheres that models the three-dimensional attributes of a molecule that are most important for binding of a ligand to its receptor. Each sphere represents the region in space that should be occupied by a chemical functionality capable of the kind of interaction specified by the feature type, for optimal drug-receptor interaction.[250] Each chemical functionality is assigned a weight, which specifies its relative importance in the hypothesis; a large weight indicates that the feature is more important, or plays a greater role in exerting biological activity, than the other features of the hypothesis.[295] The version of the Catalyst program used in this research[291] assigned the same weight to each feature during the automated hypothesis generation.

A Catalyst hypothesis can be constructed by using positioning data from x-ray crystallographic, NMR experiments or modelled receptors, or from structure activity data. Chemical feature-based pharmacophores derived from structure activity data can be a successful drug design tool provided two assumptions about the data can be made. The first of these is that all of the active compounds in the set bind to the same receptor, and in roughly the same fashion.[250] The second assumption is that compounds that have more binding interactions with the receptor or binding site are more active than those compounds that have fewer interactions. The corollary to this second assumption is that less active compounds tend to be smaller and possess fewer chemical features than more active compounds.[250] However, in any data set there may be some large and feature-rich molecules that are not very active, possibly due to a conflicting steric interaction between the receptor and a portion of the molecule that may not be actually mapped to the pharmacophore. These can be identified by the use of exclusion spheres that describe receptor constraints, added manually to the hypothesis. In the event that no information about the receptor structure is available, such compounds are better left out of the analysis as Catalyst overestimates their activities.[250]

4.2.1 Training set selection

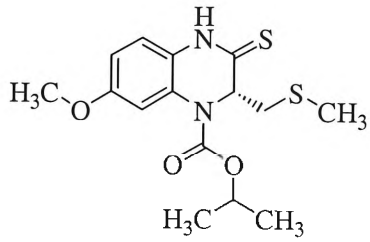
The selection of compounds in the training set is essential to a successful application of the pharmacophore analysis, as the hypotheses are derived entirely from the information in the training set.[296] Structurally diverse compounds, containing clear and concise information, are required; each compound should be selected to provide as much unique

information in terms of structure or activity as is possible.[250] Structural redundancy needs to be avoided as this biases the statistical calculations used to correlate predicted and measured activities during the analysis.

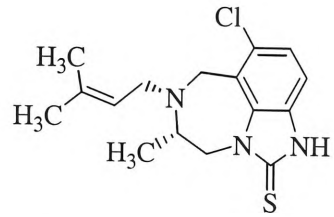
The training set ideally needs to contain at least sixteen compounds to be statistically significant.[297] Activity values should span four orders of magnitude, and each order of magnitude should be represented by at least three compounds. If there are two structurally similar compounds, they are both retained only if their activities differ by one order of magnitude, otherwise only the most active of the two compounds should be selected. If two compounds have similar activities, within one order of magnitude, they must be structurally distinct for both to be included in the analysis, or else only the most active of the two should be selected. Compounds known to be inactive due to steric obstruction of the receptor should be excluded, though this is not a consideration in this particular study.

The training set for these studies was derived from the structures of inhibitors previously crystallised in RT-inhibitor complexes (Appendix 7.3 X-ray crystallographic structures of reverse transcriptase). Not all inhibitor structures were used in the training set, as activity data could not be obtained for all of the inhibitors. It was impossible to obtain reliable and consistent activity data for all of the structures, as screens, assays, cell lines and experimental conditions used for inhibitor testing varied between research groups, as discussed in Section 3.8.2. However, for these preliminary studies of integrated drug design techniques, the judicious use of consistent and comparable activity data was deemed satisfactory. IC_{50} values used were obtained preferably from the same publication that reported the crystallographic data. In cases where this was not possible, IC_{50} values were obtained from publications reporting the synthesis of the inhibitors.

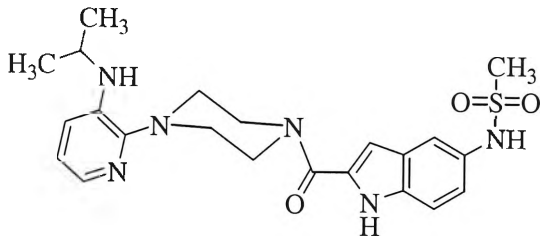
Inhibitors were identified by the corresponding PDB file code of the RT-inhibitor complex, and those used in the training set are illustrated in Figure 4.3.



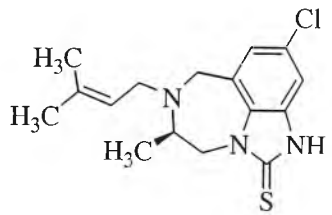
bqm - HBY097 (**111**)



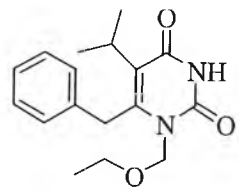
hnv - TIBO R86183 (**11**)



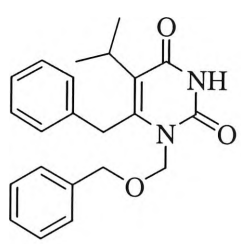
klm - BHAP U90152 (**113**)



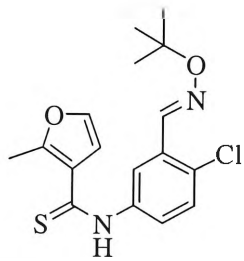
rev - TIBO R82913 (**114**)



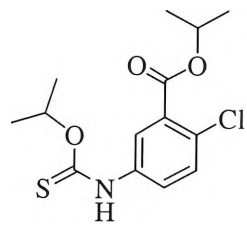
rt1 - MKC-442 (**12**)



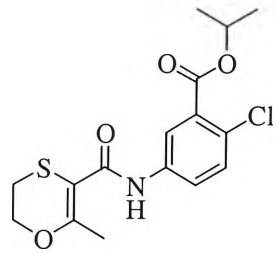
rt2 - TNK-651 (**115**)



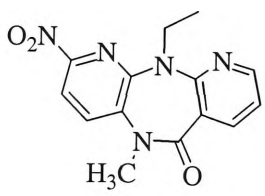
rt5 - UC10 (**118**)



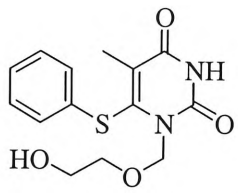
rt6 - UC38 (**119**)



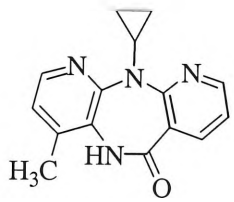
rt7 - UC84 (**120**)



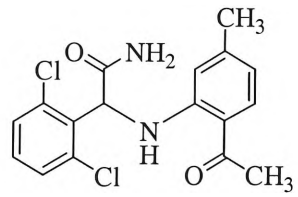
rth - 1051U91 (**116**)



rti - HEPT (**121**)



vrt - nevirapine (**8**)



vru - Cl-α-APA (**122**)

Figure 4.3 Structures of the HIV-1 RT pharmacophore training set.

Six of the inhibitor structures were excluded from the pharmacophore analysis for several different reasons. The 1bqn inhibitor, HBY097 (**111**), was excluded as it was crystallised in a mutant enzyme, and the same inhibitor was available in the 1bqm RT-inhibitor complex. 1hni and 1rt4 were excluded, as no activity data could be determined for the inhibitors Br-α-APA R 95845 (**112**) and UC781 (**117**), respectively. 1rt3 and 1uwb were excluded as they had the same inhibitor as 1rth, 1051U91 (**116**), and 1hnv, 8-Cl TIBO (**11**) respectively, and both were crystallised in a mutant RT enzymes. 1tvr was not included in the training set as the same inhibitor was available in the 1rev complex that had a higher resolution, 2.6 Å compared to the 3.0 Å of 1tvr. The rationale of excluding structures crystallised in mutant enzymes from the selection of

compounds for the training set was in consideration of the subsequent structure-based drug design methodology. The training set of thirteen compounds (Figure 4.3) fulfilled none of the fundamental requirements for a pharmacophore training set, and as such the results could only be used in a qualitative manner, and as a model for the proposed method of integrated drug design.

4.2.2 Conformational analysis

In Catalyst, the flexibility of a molecule, and thus its ability to conform to a receptor or active site, is modelled by considering each compound as a collection of conformers.[246] The conformers represent a broad coverage of conformational space available within a reasonable and specified energy range, and are selected so that differences in interfunction distances are maximised. A maximum of 256 conformations can be generated for each ligand.[297] The *Best-quality* conformation generation method is recommended for hypothesis generation as it provides the most comprehensive treatment of flexible ring systems, as well as the greatest conformational coverage within Catalyst.[298] The fundamental goal of this approach is to find, from the test set compounds, the simplest set of chemical functions that correlate best with the observed activity.[246]

The inhibitor crystal conformations were imported into Catalyst and categorised as Single RT ligands, and no conformations were generated for this training set. A second set of these thirteen inhibitors was created, and from this set, conformational models were generated; the original crystal conformation was included with these models. This second set of inhibitors was categorised as Multiple RT ligands. Table 4.1 details the number of conformers generated for each ligand in the Multiple RT series. This gives a qualitative measure of the flexibility of a molecule.

inhibitor	no. conformations	activity (nM)
bqm	37	80
hnv	39	4.3
klm	82	1100
rev	44	1.5
rt1	21	12
rt2	28	6
rt5	68	200
rt6	51	150
rt7	40	280
rth	10	400
rti	58	66000
vrt	5	84
vru	142	5

Table 4.1 Inhibitor crystal structures and the number of conformations generated using the Best-quality method. The IC_{50} inhibition data to be utilised for the pharmacophore development is also shown.

4.2.3 Hypothesis generation

Catalyst hypotheses are generated based on the assumption of Occam's razor, a principle which proposes that a theory that provides a unified, common treatment of different phenomena is preferable to a theory that treats each phenomenon as a separate and independent problem.[299] In terms of pharmacophore generation, Occam's razor signifies that between otherwise equivalent alternatives, the simplest model is best.[297]

Catalyst generates its pharmacophores in three basic phases. In the initial constructive phase, Catalyst identifies hypotheses that are common among the active compounds in the training set, before removing hypotheses that are common among the inactive compounds in the subtractive phase.[296] In the final optimisation stage, modifications are made to the hypotheses to improve their score, which provides a measure of the quality of the hypothesis.

A maximum number of five different types of descriptors or features are allowed for consideration in the hypothesis generation,[297] and can be chosen from hydrogen bond donors and acceptors, aliphatic and aromatic hydrophobes, and positive and negative ionisable groups.[300] Hydrogen bond donors and acceptors are composed of a heavy atom on the ligand and a vector projecting to the point of a complementary receptor atom. This provides directionality and further discrimination in the search query. Aromatic hydrophobes are also specified as ring aromatics, which more specifically

describes the π-π interaction between aromatic features, by restricting the plane of the aromatic system instead of just defining the centre point of the aromatic ring system. A vector is also projected to the centre point of the interacting aromatic system of the receptor in this feature. Function mapping of the inhibitors can identify features that are sterically accessible to the receptor for intermolecular interactions; the mapping of some features may depend on the conformation of the ligand. Five features were identified from the test set ligands; hydrogen bond acceptor, hydrogen bond donor, hydrophobic aliphatic, positive ionisable, and ring aromatic groups, and these were selected, unmodified, for use in the pharmacophore generation. A positive ionisable feature represents a functional group, such as a tertiary amine, that can be protonated under physiological pH. The ring aromatic function was utilised instead of the hydrophobic aromatic, as it was a more specifically defined functionality, containing more information about the location of the interacting π system on the protein.

Catalyst pharmacophores for the Single RT and Multiple RT training sets were generated in HypoGen using standard parameters,[297,298] with two exceptions. The *Minimum Total Features* parameter was increased from the standard value of 1 to 3 for the Single RT, and to 3 and 4 for the Multiple RT training sets. A *Minimum Total Features* value of 4 and 5 for the Single RT and Multiple RT training sets, respectively, failed due to unfavorable cost values. Increasing this parameter changes the hypothesis space created in the constructive and subtractive phase of the pharmacophore generation and forces Catalyst to search for hypotheses with more features.[301] Hypotheses with more features are more likely to be further discriminating and specific for that particular receptor or biological target.

The second parameter that was modified was the *forceAbsoluteSterochemistry* option that is generally set to 0, indicating that the *Mirror Image* function is on. This function considers the mirror image of the ligands during the hypothesis analyses to obtain greater conformational coverage at less computational expense.[301] However, the mirror image of ligands in which an absolute chiral center is assigned are not utilised in the hypothesis analysis. As the structures of the ligands in the training sets had been obtained from crystallographic structures, and the conformations of those in the Single RT set are actually those observed in the binding pocket, it was determined that the *Mirror Image* function was unnecessary, and might dilute the specificity of the resulting hypothesis. Pharmacophores of the Single RT and Multiple RT training sets were

therefore generated with the *Mirror Image* function on and off and were designated Singlepharm and Multipharm, and Chiralsinglepharm and Chiralmultipharm, respectively.

4.2.4 Hypothesis analysis and selection

To design a pharmacophore, Catalyst generates up to ten unique hypotheses according to how well the training set fits the descriptors used. Cost analyses of the hypotheses are calculated to determine the ranking of the pharmacophores, and mapping patterns of the compounds help to decide which hypothesis best correlates the data. The cost analyses of all the hypotheses generated are presented in Table 4.2.

At the start of an automated hypothesis generation HypoGen calculates the cost of two theoretical hypotheses representing the upper and lower bounds of the training set.[296,297] A fixed hypothesis represents the ideal hypothesis cost, a value assigned to the simplest possible hypothesis that fits the data precisely. The null hypothesis assumes that there is no statistically significant structure within the data, and that the experimental activities are randomly distributed about their mean. A fixed cost and a null hypothesis cost are generated to give an initial indication of the probability of finding useful pharmacophores. They are expressed in bits, and the difference between the two is a measure of the probability of finding statistically valid and useful pharmacophore models; the greater the difference, the higher the probability.[297]

The null cost of the pharmacophore runs of Table 4.2 were all the same, while the fixed cost varied slightly with changes in the *Minimum Total Features* parameter and the use of the multiple conformations of the inhibitor crystal structures in the Multiple RT training sets. The *Mirror Image* parameter had no affect on the fixed and null hypotheses, and only influenced the ten individual hypotheses generated in each pharmacophore run. The average difference between the fixed and null hypotheses in the pharmacophores generated from the Single RT training set, where only the conformations from the crystal structures were used, was 35.1. This average difference in the pharmacophores of the Multiple RT training set, where conformational analyses of the ligands were performed, was 30.0. This appeared to indicate that the hypotheses generated from the Single RT training set were more likely to be statistically valid and a more faithful correlation of the training set ligand structures and inhibition data than for those pharmacophores generated from the Multiple RT training set.

		Single RT				Multiple RT					
		singlepharm		chiralsinglepharm		multipharm			chiralmultipharm		
	Mirror Image	on	on	off	off	on	on	on	off	off	off
	Min. Total features	1	3	1	3	1	3	4	1	3	4
	Hypothesis features	AHR APR	AHR APR	AHR APR	AHR APR	AHHR	AHHR AHPR	AHHR	AHHR	AHHR AHPR	AHHR AHPR
	Fixed cost	50.0	49.8	50.0	49.8	55.2	55.2	54.7	55.2	55.2	54.7
	Null cost	85.0	85.0	85.0	85.0	85.0	85.0	85.0	85.0	85.0	85.0
	Entropy	5.1	5.0	5.1	5.0	10.3	10.3	9.8	10.3	10.3	9.8
Relative hypothesis cost	hypothesis 1	21.0	20.7	21.1 §	21.3	20.7 §	17.7 *	22.5 §	19.1	18.9	22.2 §
	hypothesis 2	20.5	20.0	20.9	21.1	16.3	15.3	12.5	15.4 §	11.5	18.7 §
	hypothesis 3	19.3	18.5	19.4	20.6	14.1	15.1 §	11.0	12.7	9.6	17.5 §
	hypothesis 4	18.8	17.1	18.2	20.4 *	11.9	14.7 *^	10.8 §	12.0	9.4	10.5 §
	hypothesis 5	16.4	16.6	17.1	17.4	11.8 #§	13.8 *^	10.4	6.6 #§	9.0 *^	10.4
	hypothesis 6	16.3 *	16.6	15.8	17.1	11.3	12.3	10.1	6.0 *	8.6 §	9.7 §
	hypothesis 7	13.4	16.3 *	13.6	14.7	10.8	11.2	9.3	5.4 #§	8.6	8.6
	hypothesis 8	11.4 §	12.6	12.6 *	12.3 *	7.5	9.7	8.4	5.3 #§	8.1 #§	8.4
	hypothesis 9	7.2 ^	11.0 ^	12.2	8.0 ^	7.5	9.0 §	7.3 §	5.3 #§	5.7 #§	8.2 ^
	hypothesis 10	6.5 ^	6.3 ^	9.9 ^	4.6 ^	7.4	8.1 #§	6.7	4.9	4.8 §	7.7

Table 4.2 Summary of the cost analyses of Single RT and Multiple RT pharmacophores.

Hypothesis features summarises the number and type of features in the ten returned pharmacophores. The first set was the most common.
A – hydrogen bond acceptor, P - positive ionisable, H - hydrophobic aliphatic, R – hydrophobic ring aromatic
Relative cost is the difference between the cost of the null hypothesis and the cost of the individual hypothesis.
* - one or both of the 2 most active compounds did not map well onto the hypothesis.
§ - the three top active compounds mapped well onto the hypothesis.
# - hypothesis has a hydrophobic aliphatic feature missing from the first described set of features.
^ - hypothesis corresponds to the second set of features.

Initially this observation appeared erroneous when one considered that the Single RT training set contained only 13 conformations of ligands compared to the 625 ligand conformations of the Multiple RT training set. However, the fact that the 13 singular conformations of the ligands of the Single RT training set were derived from crystallographic structures, and were most likely to be, or very closely approximate, the active conformation of the inhibitor, meant that more accurate information was used to calculate those hypotheses. The use of the multiple conformations in the Multiple RT training set thus appeared to dilute the pharmacophoric analysis, returning less statistically robust hypotheses.

Each of ten hypotheses generated is also accorded a cost value in HypoGen. The relative cost is the difference between the cost of each of these hypotheses and the cost of the null hypothesis, and relates to its probability of representing a true correlation of the data.[297] A difference of more than 60 bits means that there is a greater than 90 % chance that the hypothesis is a true correlation, a difference of 40-60 bits equates to a probability of 75-90 %, and for any value below 40 the percentage drops to below 50 %. Under these last conditions it is unlikely that the models will be predictive, or represent a statistically valid correlation of the training set data.

The relative costs of the hypotheses of Table 4.2 revealed that all cost analyses fell well outside of the range considered functional for pharmacophore analysis. Furthermore, nearly all of the returned hypotheses had relative costs less than 20, a situation in which the model, as a rule, has little chance of succeeding, and it is generally advised to reconsider the training set.[297] However, this was not considered overly significant in this study where the primary objective was to assess the integration of ligand- and structure-based drug design tools. Moreover, unlike most pharmacophore studies, all of these ligands were definitively known to bind at this one biological target, and in the case of the Single RT training set, the actual binding mode of the ligands was available for the pharmacophore analysis.

There was surprisingly little difference between highest ranking hypotheses of the Single RT and Multiple RT training sets, with average relative costs of 21.0 and 20.2 respectively. However, it was noted that the hypotheses generated from the Single RT training set displayed a gradual decrease in the relative cost from the first hypothesis to the last hypothesis, whereas the hypotheses of the Multiple RT training set often

displayed a large difference between the first and second hypotheses. This was most notable in the multipharm pharmacophore where the *Minimum Total Features* parameter was set to 4; the difference between the first and second hypotheses was 10.1 bits.

Very little difference was also seen between hypotheses generated with and without the *Mirror Image* function. In the Single RT training set, for the highest ranking pharmacophore, the average relative cost was 20.8 with the Mirror Image function on compared to 21.3 without the same function, and similarly the average of all ten hypotheses was 15.3 and 15.9 with the function on and off respectively. In the Multiple RT training set, the reverse effect was observed with the average relative cost slightly higher with the *Mirror Image* function on rather than off, 20.3 compared to 20.1 respectively for the highest ranking hypothesis, and 11.8 compared to 10.3 for all ten generated hypotheses.

Modification of the *Minimum Total Features* parameter appeared not to have a significant or compelling influence on the relative costs of the returned hypotheses, despite the observation that the highest ranked hypothesis of 22.5 was generated in the multipharm pharmacophore where the *Minimum Total Features* was set at 4. In the case of the chiralsinglepharm pharmacophore, increasing this parameter from 1 to 3 improved the costing of the hypotheses, while in the case of the singlepharm pharmacophore, it decreased the costs of the highest ranking hypotheses. In the multipharm and chiralmultipharm pharmacophores the modification of the *Minimum Total Features* parameter from 1 to 3 to 4, first decreased and then increased the cost of the highest ranking hypothesis. The pattern observed for the overall average of the ten returned hypotheses in each of the pharmacophore runs, for both the multipharm and chiralmultipharm pharmacophores, was first an increase and then a decrease in the average hypotheses cost.

Another gauge of the validity of the hypotheses is the entropy, or configuration cost, a parameter describing the complexity of the hypothesis space HypoGen must analyse, equal to the exponent to the base 2 of the number of models Catalyst will attempt to optimise during the hypothesis run.[296,297] Entropy values up to 17 are reasonable for the optimisation of initial hypotheses. Values greater than 17 generally indicate that the training set ligands may be too flexible to allow a correct space coverage during conformational analysis, and not all pharmacophore models may be considered.[296] The

entropy in the Single and Multiple RT training sets were not affected by the *Mirror Image* function, but were affected by the *Minimum Total Features* parameter, generally decreasing as this parameter was increased. The average entropy, or configuration cost, for pharmacophores derived from the Single RT training set was 5.042, and twice as much for pharmacophores derived from the Multiple RT training set, which had average entropy values of 10.2.

Although the cost elements are a significant consideration in the selection of a final hypothesis from the ten returned hypotheses of a pharmacophore run, other factors need to be taken into account to fully exploit the Catalyst generation process.[296] The main assumption made by HypoGen is that an active ligand should map more features than an inactive ligand. Consequently the mapping information of the ligands, should be used to assess the quality of the hypotheses.[296] In particular, the two most active ligands should map all features of a pharmacophore, as the initial hypotheses are derived from the top two ligands of the training set.

The pharmacophores derived from the Multiple RT training set were generally four-feature hypotheses, compared to the Single RT training sets that generated only three-feature hypotheses. The Multiple RT pharmacophores usually consisted of a hydrogen bond acceptor, two hydrophobic aliphatic functionalities, and one ring aromatic system. Most of the returned hypotheses were fairly similar on a gross level, with small differences predominately seen in the spatial positioning of features, the angle of the plane of the aromatic ring system, the direction of the vector describing the location of the interacting π system on the protein, and the positioning and directionality of the heavy atom and vector of the hydrogen bond acceptor. In the multipharm pharmacophore where the *Minimum Total Features* parameter was set at 3, and the chiralmultipharm pharmacophores where this parameter was set at 3 or 4, hypotheses were generated that included a positive ionisable group instead of the second hydrophobic aliphatic group. These hypotheses had fairly poor relative costs, the highest being only 14.676, and most probably were only generated because the two most active ligands, rev and hnv, each included a positive ionisable function, which was absent from all other ligands. The similarity in the structures and biological activities of rev and hnv would customarily have meant that hnv, with the lower IC_{50} value of 4.3 nM, would have been excluded from the training set as redundant data. However, given that the size of the training set was only thirteen ligands, the evaluation was made that

its inclusion was necessary and an allowance could be made for the bias that its contribution to the pharmacophore generation would engender. This compromise was reached by examining hypotheses in which the top three compounds mapped, and not just considering hypotheses in which the top two ligands mapped. These hypotheses are marked by § in Table 4.2, and take into account the structure and activity of vru. They occur notably more often in the pharmacophores generated from the Multiple RT training set, in particular when the *Mirror Image* function is off.

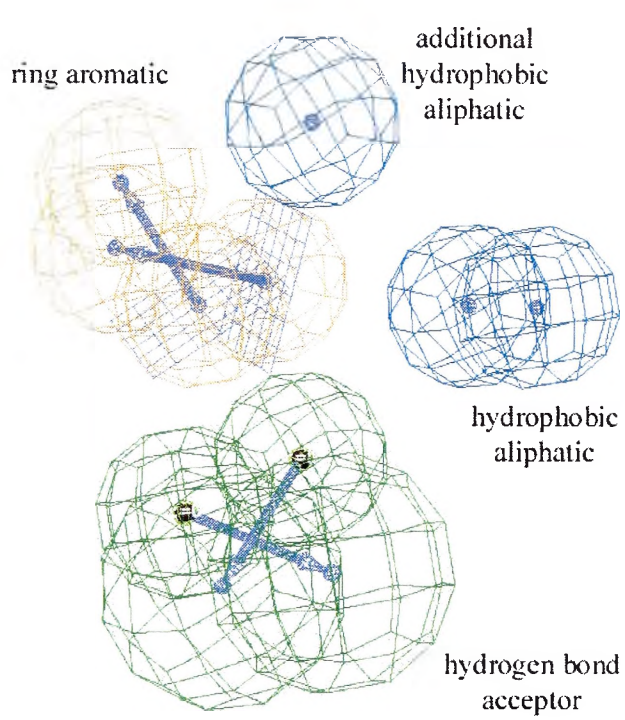


Figure 4.4 Overlay of a typical three-feature hypothesis generated from the Single RT training set and a four-feature hypothesis generated from the Multiple RT training set. The wire-cage spheres represent the pharmacophore features, the small blue dot within the blue hydrophobic feature signifies the centre of the sphere, or the point that the ligand functionality maps in the feature. The blue square grid and blue arrow of the beige ring aromatic feature represents the mapped plane of the aromatic ring and the projected interaction with the corresponding aromatic amino acid of the receptor. The blue arrow within the green hydrogen bond acceptor feature, sometimes seen as a green cone) represents the projected hydrogen bond donor of the receptor at the blue dot.

The returned hypotheses from the Single RT training set all contained only three-feature pharmacophores, most of which were composed of a hydrogen bond acceptor, a hydrophobic aliphatic functionality, and a ring aromatic system. Figure 4.4 illustrates the difference between the highest ranking multipharm pharmacophore (*Minimum Total Features* parameter set at 4), a four-feature pharmacophore, and the highest ranking hypothesis from the singlepharm pharmacophore (*Minimum Total Features* parameter set at 1). The last one or two hypotheses generated in each of the Single RT training set pharmacophore runs were hypotheses in which the hydrophobic aliphatic group was replaced by a positive ionisable group. This occurrence was likewise explained by the biasing of the pharmacophore by the inclusion of rev and hnv, two similarly structured and active ligands with inherent positive ionisable

functionalities. Only two of the returned hypotheses from the Single RT training set fulfilled the requirements of the compromised examination of hypotheses to consider the top three most active compounds; the highest ranking hypothesis from the chiralsinglepharm pharmacophore run with 1 set in *Minimum Total Features* parameter, and the eighth highest ranking hypothesis from the singlepharm pharmacophore run with the Minimum Total Features parameter set also at 1.

One point of interest noted in the examination of the returned hypotheses of Table 4.2 was the lack of an expected second aromatic ring system. Literature describing the non-nucleoside inhibitor binding pocket (NNIBP), and the 'butterfly' conformation adopted by many inhibitors bound in this pocket, refers often to ring aromatic π-stacking interactions between the hydrophobic 'wings' of the butterfly and aromatic residues of the pocket, in particular Tyr181, Tyr188, Phe227, and Trp229[30,215] (see Section 3.7.1 Inhibitor binding in the NNIBP and Figure 3.49). These pharmacophores indicate that the hydrophobic 'wings' of the butterfly are composed of one hydrophobic aromatic ring system, and one hydrophobic aliphatic functionality. In Figure 4.4 the hydrophobic aliphatic and aromatic ring system common to the two overlaid pharmacophores are opposite each other, as if to represent the wings of the butterfly conformation, with the hydrogen bond acceptor serving as the body of the butterfly.[30]

With the relative costs and mapping information of the returned pharmacophores considered, several hypotheses were selected for further analysis and validation, and for incorporation into a more comprehensively descriptive pharmacophore. From those pharmacophores generated with the Single RT training set, the highest ranking hypothesis from the singlepharm (*Minimum Total Features* set at 1) and chiralsinglepharm (*Minimum Total Features* set at 3) pharmacophores were selected. Also selected was the ninth ranking hypothesis from the singlepharm (*Minimum Total Features* set at 3) pharmacophore, that had a relative cost of 10.967; the highest ranking hypothesis that also contained the positive ionisable functionality. From the Multiple RT training set, the highest ranking hypothesis of the multipharm (*Minimum Total Features* set at 4) pharmacophore, and the fourth highest cost hypothesis from the multipharm (*Minimum Total Features* set at 3) pharmacophore, containing a positive ionisable group, were selected. These choices were subjective, based loosely on the relative costs and mapping information available, and may not be ideal, but rather they

provide an example of the techniques and validation processes that one might choose to explore for greater application of pharmacophoric ligand-based drug design strategies.

4.2.4.1 Merged hypotheses

Depending on the ligands of the training set submitted to HypoGen, the generation of Catalyst pharmacophores can produce varying results in which a particular functionality may only appear in some hypotheses. The incorporation of a particular feature into another hypothesis either without that feature, or lacking that feature in a specific position relative to other common features, can be accomplished by manually adding and positioning the feature, or by merging hypotheses.[291,298] The amalgamation of two hypotheses has the advantage of allowing the construction of more comprehensively descriptive and specific pharmacophores. However, this can also exclude compounds, during a database search, that may bind just as well in the receptor site, but may not have all of the features of the pharmacophore, and are consequently not considered at the outset of the search. This depends on the particular search algorithm used, and if partial-match queries, in which not all the features need to be mapped by a ligand in order for it to be a 'hit', are available for database mining.[280,302] Notwithstanding this particular obstacle, four hypotheses were selected to be merged, and while they may be unsuitable and ineffective for database mining, they are likely to be valuable in *de novo* ligand design, where no penalties are made for more complicated or feature-rich pharmacophores.

The highest ranking hypothesis from the singlepharm (*Minimum Total Features* set at 1) pharmacophore and the ninth ranking hypothesis from the singlepharm (*Minimum Total Features* set at 3) pharmacophore were chosen to be merged (Figure 4.5). The expectation of this choice was to construct a pharmacophore that incorporated the high ranking spatial placement of the hydrogen bond acceptor, hydrophobic aliphatic and aromatic ring system of the higher ranked hypothesis with the positive ionisable functionality of the other hypothesis. This was achieved using the Merge Hypothesis/Features tool in Catalyst. In this tool, a *Distance Tolerance* parameter specifies the maximum separation between a pair of common features from two hypotheses for them to be merged into one feature which can be considered the average of the pair of common features.[297] This parameter was increased from the standard 0.1 Å to 0.5 Å, to allow complete merging of the two singlepharm pharmacophores, into a

four feature pharmacophore with hydrogen bond acceptor, hydrophobic aliphatic, ring aromatic and positive ionisable functions.

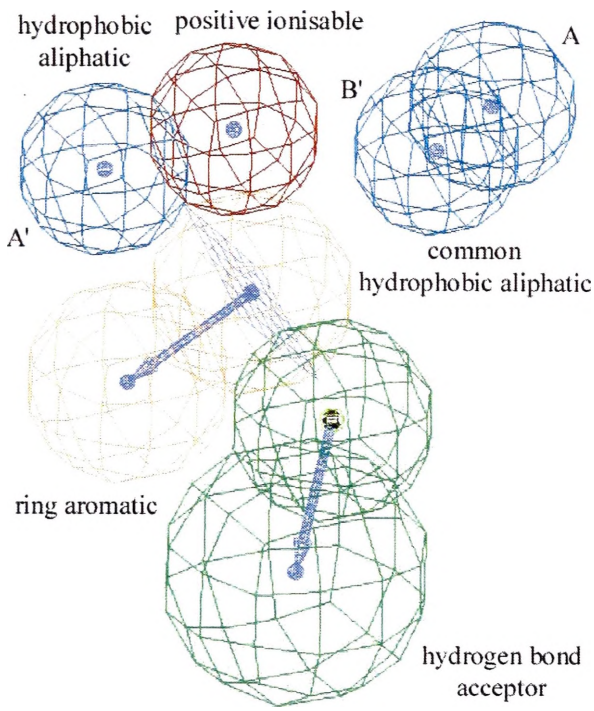


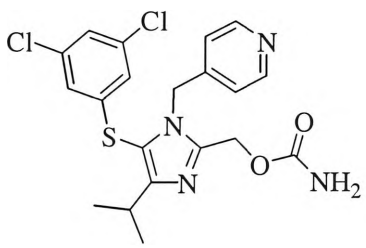
Figure 4.5 Merged multipharm pharmacophores. The A hydrophobic aliphatic function was from the hypothesis including the positive ionisable group. The A' and B' aliphatic functions were from the higher ranking hypothesis.

This *Distance Tolerance* parameter had to be increased further to 0.8 Å in the merging of the highest ranking hypothesis of the multipharm (*Minimum Total Features* set at 4) pharmacophore and the fourth hypothesis from the multipharm (*Minimum Total Features* set at 3) pharmacophore, containing a positive ionisable group. In Catalyst 4.0, the version in which the greater part of the research was realised, the two hydrophobic aliphatic features needed to be renamed. This was because the merging process first overlaid the A hydrophobic aliphatic of the lower ranking hypothesis (Figure 4.5) to the A' hydrophobic aliphatic of the higher ranking hypothesis, rather than the B' function as preferred. A *Distance Tolerance* value of 0.8 Å was just sufficient to ensure that the hydrogen bond acceptor and ring aromatic functions were merged properly, however it was insufficient to merge the two common hydrophobic aliphatic functions, A and B'. The A hydrophobic aliphatic function was ultimately chosen for the merged hypothesis as it was placed further away from the other functionalities, and would hopefully 'open up' the pharmacophore rather than constructing it so densely packed with features in the middle, and the B' hydrophobic aliphatic was removed. In retrospect, it may have been better to use the B hydrophobic aliphatic function, as this came from the higher ranking hypothesis, or merge the two common aliphatic features within the pharmacophore to achieve an average representation. The final pharmacophore contained a hydrogen bond acceptor, two hydrophobic aliphatic functionalities, an aromatic ring and a positive ionisable function that lay in contact with one of the hydrophobic aliphatic spheres. The two resulting

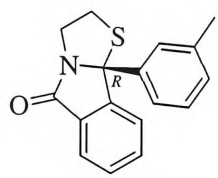
merged hypotheses were referred to as the merged singlepharm and merged multipharm hypotheses

4.2.5 Hypothesis validation

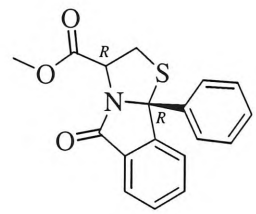
Validation of a hypothesis is achieved be determining how well activities of non-training set compounds are estimated, visually and mathematically.[250] The best hypotheses will estimate the activities of the training set compounds correctly to within one order of magnitude, as well as passing standard statistical tests for non-randomness.[250] Although numerous analogues of the ligands used in the pharmacophore generation and illustrated in Figure 4.2 could have been used for this validation process, the premise of using inhibitors from crystallised RT-inhibitor complexes was continued in the selection of ligands for the test set. The ligands from two recent RT crystallographic references,[303,304] not used in the training set, were obtained and processed in the same fashion as those used in the pharmacophore generation (Section 4.2.2 Conformational analysis). These crystallographic structures were tabulated in Appendix 7.3.1. The three ligand structures, illustrated in Figure 4.6, were identified by their PDB file code, and categorised as Single or Multiple test sets, dependent on whether they contained the ligands in their original crystal conformations or multiple conformations generated in the Best-quality mode of the catConf module.



ep4 - S-1153 (**123**)
IC_{50} = 450 nM



c0t - BM +21.1326 (**124**)
IC_{50} = 4100 nM



c0u - BM +50.0934 (**125**)
IC_{50} = 220 nM

Figure 4.6 Structures and names of inhibitors from recent x-ray crystallographic RT-inhibitor complexes to be used as ligands in the test or validation set.

Once more, the difficulty in obtaining reliable and consistent activity data among these ligands, and among those used in the pharmacophore generation, caused the validation process to be rather problematic, with results having only qualitative value. The IC_{50} values for the ligands ep4,[305] c0t[306] and c0u were 450, 4100 and 220 nM respectively, and were obtained from different sources. Realistically, comparisons could only be

made between c0t and c0u, as they had the same testing procedures, and even then they could only be qualitatively assessed as the pharmacophores to which they were mapped were generated with uncertain inhibition data from varying testing protocols.

The test sets were mapped or overlaid onto the selected pharmacophores of Table 4.3 with the Compare/Fit tool to calculate how well they correlated geometrically and/or functionally to the hypothesis.[297]

		singlepharm min. feature - 1 hypothesis - 1 21.0		chiralsinglepharm min. feature - 3 hypothesis - 1 21.3		multipharm min. feature - 4 hypothesis - 1 22.5		multipharm min. feature - 3 hypothesis - 4 14.7	
	rms	1.416		1.427		0.867		1.448	
	correl.	0.833		0.827		0.954		0.828	
		fit	estimate	fit	estimate	fit	estimate	fit	estimate
Single	ep4	4.7	89	4.6	90	5.6	5100	4.7	460
	c0t	2.8	6200	3.9	470	5.4	8300	4.5	760
	c0u	4.8	68	3.8	570	5.4	9600	4.6	510
Multiple	ep4	6.8	1	6.7	1	7.8	39	6.9	3
	c0t	4.7	82	5.3	20	6.9	300	4.7	400
	c0u	4.8	57	4.3	200	6.0	2400	4.9	260

Table 4.3 Summary of the fit and estimate values for the test set ligands, in the conformation found in the RT-inhibitor crystal complex (Single) and in the multiple conformations derived from conformational analysis (Multiple), overlaid on four selected Catalyst hypotheses. Each pharmacophore is identified by the name, number used in the Minimum Total Features parameter, ranking of the hypothesis, and its relative cost (Refer to Table 4.2). The IC_{50} values of ep4, c0t and c0u were 450, 4100 and 220 nM, respectively.

This was done by geometrically aligning the features of the ligand to corresponding features in a hypothesis to meet specified topology, angle, torsion and chemical-function constraints. The *Fast Fit* option was used preferentially as it searches each conformation for the optimum fit; in the case of the Single ligand training set, only the one crystal conformation is used. The alternative *Best Fit* option manipulates all conformers within a specified energy threshold to minimise the distances between hypothesis features and mapped functionalities of the ligand.[297] It was not used because a *Fast Fit* method is used to fit ligands onto the pharmacophores during hypothesis generation for their evaluation, and no 'tweaking' or optimising is done. We have observed in the past that *Best Fit* can lead to overestimation of the activities (unpublished data). Both options return an optimum set of mappings representing

different alignments of ligand features to pharmacophore features. The fit value indicates the quality of the mapping and is dependent on how close the features of the ligand are to the centres of the corresponding location constraints; the higher the fit value, the better the mapping of features.[297]

The root-mean-square (rms) and correlation (correl.) values for each of the hypotheses of Table 4.3 represent the deviation of the logarithmic of the estimated activities from the logarithmic of the measured activities normalised by the log of the uncertainty, and the linear regression derived from the geometric fit index, respectively.[298] The rms indicates the quality of the prediction for the training set, and the correlation gives an assessment of the relationship between the actual and the predicted activities.

From the values of Table 4.3 it appeared that the first multipharm pharmacophore was the most successful hypothesis as it had the highest correlation between the actual and predicted, and the lowest deviation, for the training set compounds used to generate the hypothesis. However, the estimates of activity for the ligands of the Single and Multiple test sets showed that the pharmacophore erroneously calculated the c0u ligand to be less active than the c0t ligand, 9600 compared to 8300 and 2400 compared to 300 for the crystal and multiple conformations test sets, respectively. The chiralsinglepharm pharmacophore also returned the same result. Only the second multipharm pharmacophore and the singlepharm pharmacophore predicted the c0u ligand to be more active than the c0t ligand; in both cases this proved true for the ligands in the Single and Multiple test sets. However, neither of these hypotheses, for the Single or Multiple test sets, were able to closely approximate the relative ratio of activities of c0t and c0u.

As expected, the ligands of Multiple test set had lower estimated IC_{50} values than the restrictive Single test set. While some of the hypotheses were able to estimate the activity of the test set ligands within one order of magnitude, no one hypothesis was able to predict the activities of all three ligands within that same range, not even just c0t and c0u which can be more confidently compared. The closest approximation of this was the singlepharm pharmacophore that estimated c0t at 6200, compared to the actual 4100, and c0u at 68 compared to 220.

4.2.5.1 Merged hypotheses

The validity and predictive power of the merged singlepharm and merged multipharm hypotheses were analysed by seeing how well they correlated the activity of the test set ligands. The Regression Hypothesis tool in Catalyst allows a merged or manually constructed hypothesis to be used to estimate the activities of a training set by computing a linear structure-activity regression equation between the training set ligands and the hypothesis, and using this regression information to score a newly produced hypothesis.[297] The resulting hypothesis is used to estimate the activities of the ligands of training set, and these can be graphed against the true activities.

In Figure 4.7, the red Single RT training set ligands showed good correlation against the merged singlepharm pharmacophore, with an RMS in the regression hypothesis of 0.783. The same training set ligands on the merged multipharm pharmacophore of Figure 4.8 (red data series) had a slightly lower RMS regression value of 0.773. Figure 4.9 shows the overlay of the crystallographic conformation of rev, the most active ligand in the Single RT training set, on the merged singlepharm pharmacophore. All of the four features of the hypothesis were mapped and returned a Fit value of 5.700 out of a perfect fit of 8.792. The perfect fit was calculated by the sum of the average weights for the hydrogen bond donor and ring aromatic functions from the individual hypotheses, and the weight of the hydrophobic aliphatic function from one hypothesis and the positive ionisable function from the other hypothesis. This was calculated manually, and was not a value returned by Catalyst. The same crystallographic conformation of rev on the merged multipharm pharmacophore returned a Fit value of 7.668 out of 13.250, but only mapped the positive ionisable, and both hydrophobic aliphatic features; the ring aromatic feature failed to map. The perfect fit for the merged multipharm pharmacophore was calculated in a similar fashion as the merged singlepharm pharmacophore.

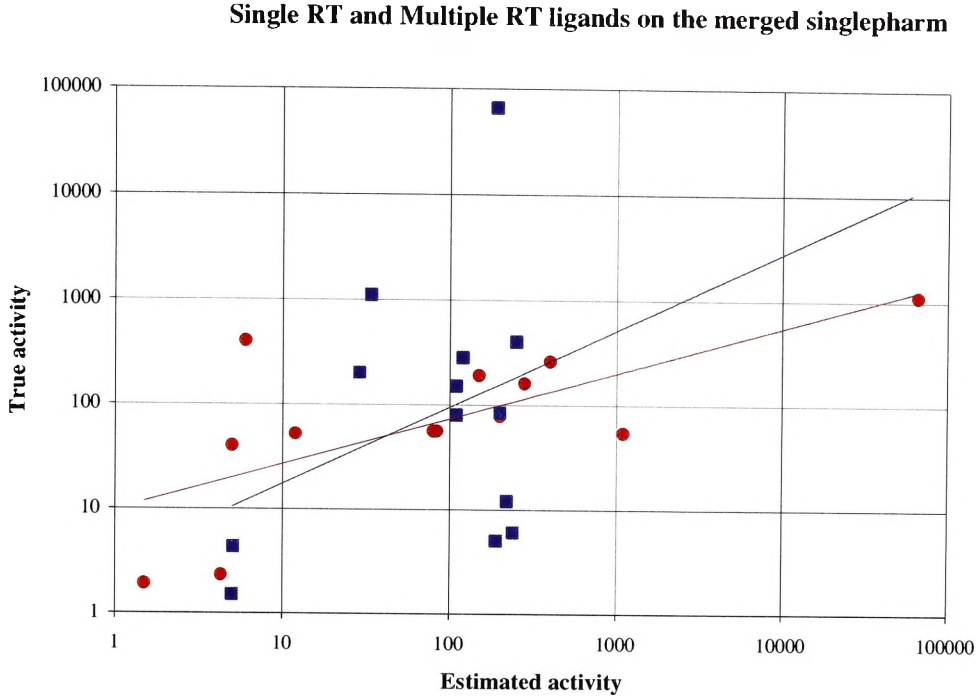


Figure 4.7 Graphed correlation of the true and estimated activities for the Single RT (red circle) and Multiple RT (blue square) training sets on the merged singlepharm pharmacophore. The line of best fit is drawn in the corresponding colour.

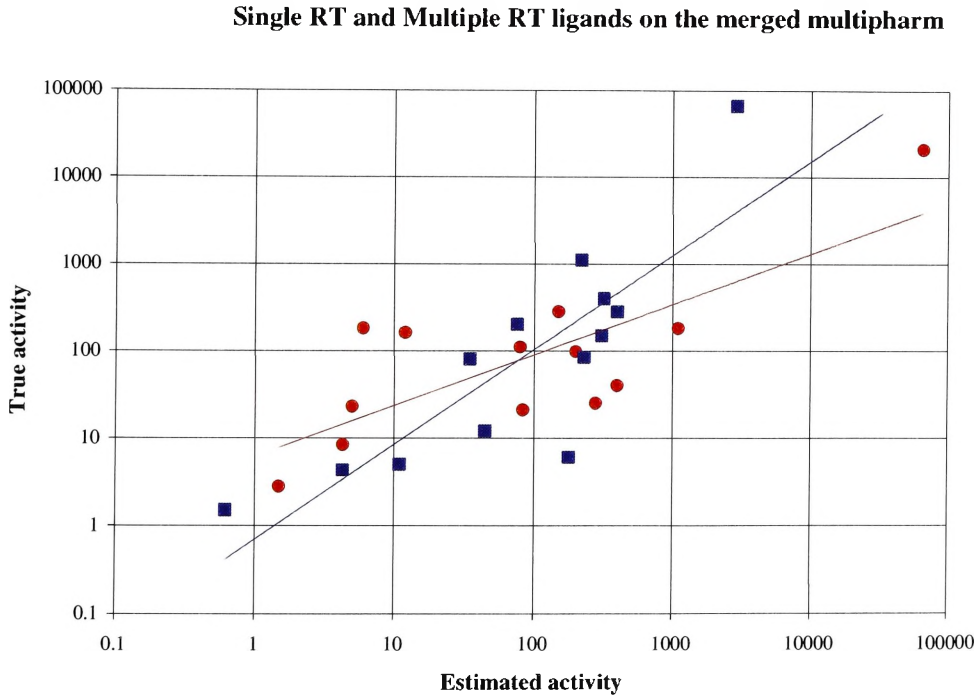


Figure 4.8 Graphed correlation of the true and estimated activities for the Single RT (red circle) and Multiple RT (blue square) training sets on the merged multipharm pharmacophore. The line of best fit is drawn in the corresponding colour

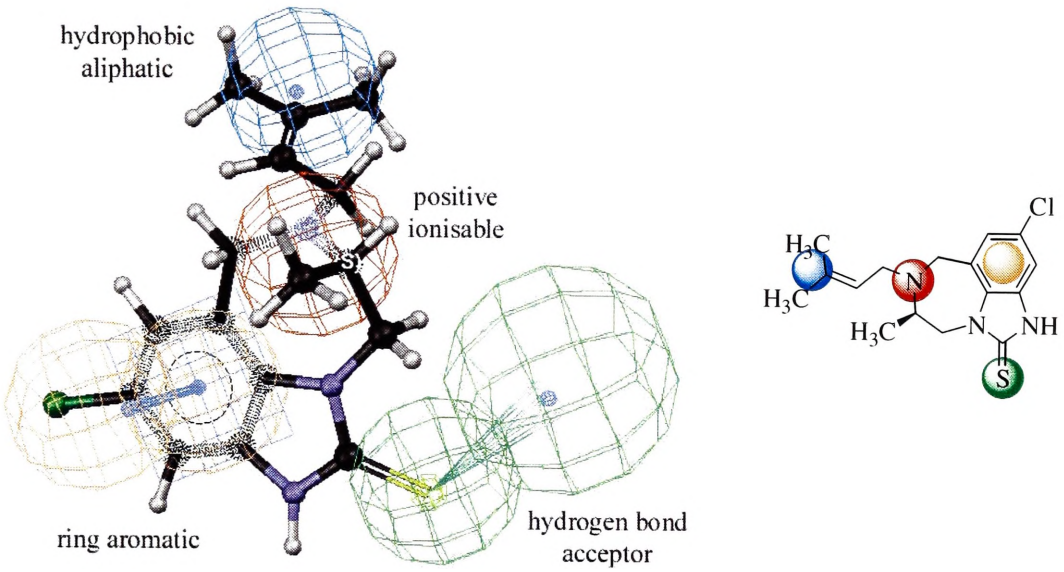


Figure 4.9 The crystallographic conformation (Single RT) of the most active ligand (rev) overlaid onto the merged singlepharm pharmacophore. (Left) The wire-cage spheres represent the pharmacophore features (see Figure 4.3 caption for full details). (Right) The coloured spheres represent the hypothesis features that have mapped to the corresponding features of the ligand.

The blue Multiple training set ligands of Figures 4.7 and 4.8 showed both the worst and best correlation values for the merged regression hypotheses, of all the training sets. In Figure 4.7 the blue data series displayed a RMS regression value of 0.357, denoting no correlation to the merged singlepharm pharmacophore. The correlation of the Multiple RT ligands on the merged multipharm pharmacophore of Figure 4.8 was much higher, at 0.836. These results showed, as may be anticipated, that the set of ligands used to generate the hypotheses of the merged pharmacophore correlated better with the merged pharmacophore compared to another set of ligands. This was even when the difference between the sets was a case of single versus multiple conformations of the same ligands. This can also be seen in the mapping of the optimum conformer of rev ligand from the Multiple RT training set, onto the merged singlepharm pharmacophore (Figure 4.10) and onto the merged singlepharm pharmacophore (Figure 4.11). The overlay of the optimum conformation of rev, the most active ligand in the Multiple RT training set, on the merged singlepharm pharmacophore in Figure 4.10, shows the same mapping as seen in Figure 4.9. All of the four features of the hypothesis were mapped and returned a Fit value of 7.058 out of likely perfect fit of 8.792. This particular conformation overlaid more optimally onto the merged singlepharm pharmacophore than the conformation actually found in the RT-inhibitor crystal complex of rev, which had a fit of 5.700 on this merged pharmacophore. The overlay of an optimum conformation of rev on the five-feature merged multipharm pharmacophore in Figure 4.11 shows the

same mapping of all features excluding the second hydrophobic aliphatic function. This appeared to coming close to mapping the methyl substituent, which was actually within the blue wire-cage sphere, but the absence of the small blue dot signified that this feature was not mapped.

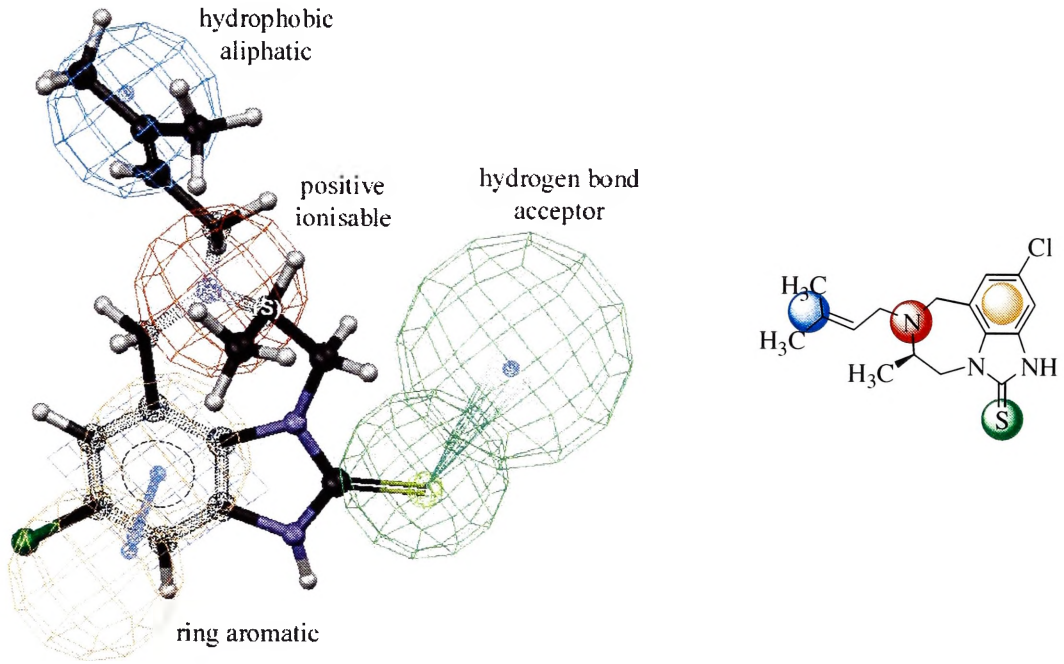


Figure 4.10 The optimum conformation (Multiple RT) of the most active ligand (rev) overlaid onto the merged singlepharm pharmacophore.

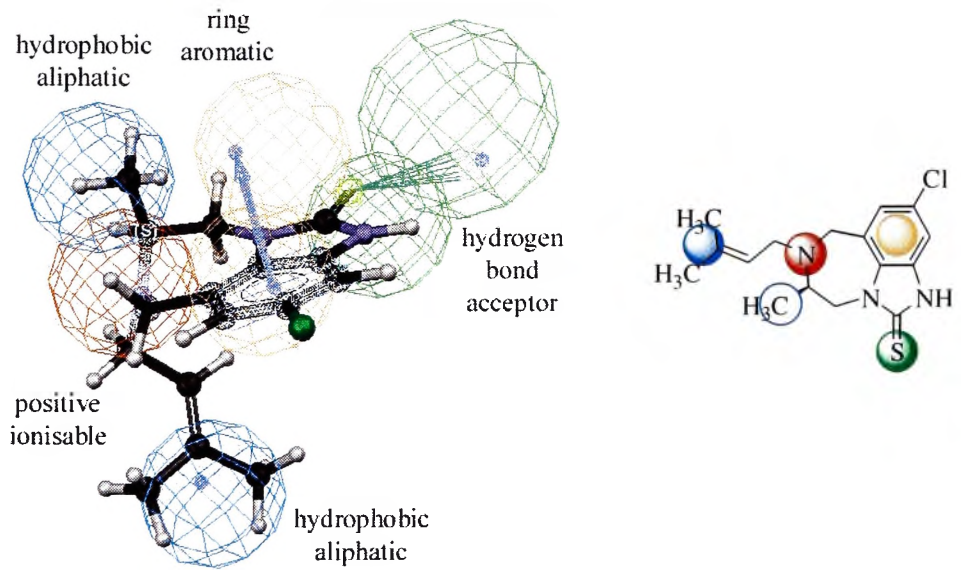


Figure 4.11 The optimum conformation (Multiple RT) of the most active ligand (rev) overlaid onto the merged multipharm pharmacophore. (Right) The blue circle outline represents the second hydrophobic function that comes close to the methyl substituent, but does not actually map to the ligand.

The overlay and Fit values of the test set ligands on the merged pharmacophores presented in Table 4.4 shed no further light on their validity or predictive power. Only the merged singlepharm pharmacophore showed a better Fit value for the c0u ligand than the c0t ligand, which had IC_{50} values of 220 nM and 4100 nM respectively. The Fit value is qualitatively inversely proportional to the IC_{50} value. This was definitively seen in ligands in the Single test set, with the conformation found in the RT-inhibitor crystal complex, but not as certainly seen in the Multiple test set where the Fit values where almost identical at 4.389 and 4.293. If one could assume the relative order of activity increased from the c0t to ep4 to c0u ligand as suggested by the experimental activities in Figure 4.6, the merged singlepharm pharmacophore was the only pharmacophore to successfully predict this order of activity among the ligands. However, these tentative findings can only be considered interesting coincidences with test sets of this size. Further validation with more ligands would be required to more thoroughly test the robustness of the merged pharmacophores.

		Fit values	
		merged singlepharm	merged multipharm
	Perfect fit	8.792	13.250
Single	ep4	4.198	6.544
Single	c0t	2.430	5.428
Single	c0u	4.323	5.369
Multiple	ep4	5.876	8.593
Multiple	c0t	4.293	6.681
Multiple	c0u	4.389	5.452

Table 4.4 Summary of the fit values for the test set ligands, in the conformation found in the RT-inhibitor crystal complex (Single) and in the multiple conformations derived from conformational analysis (Multiple), overlaid on the merged singlepharm and merged multipharm pharmacophores.

These results are not a comprehensive or complete examination of the validity of the hypotheses. Rather they establish an approach to the design of new pharmacophores

4.2.6 Database searching

Searching of the NCI database was performed with the two merged pharmacophores using Catalyst's Best Flexible Searching Tool, in Catalyst version 4.5.[307] There is no option for partial match searches, in which a search can be conducted where a match

will contain only some of the features of the pharmacophore. Returned hits must contain at least all of the features of the pharmacophore used in the search.

From an NCI database of 123 219 compounds, the merged multipharm pharmacophore identified 1162 suitable compounds. A search of these 1 162 compounds with the merged singlepharm pharmacophore revealed 984 compounds that mapped the four-feature hypothesis. Conversely, screening the 123 219 compounds of the full NCI database with the merged singlepharm pharmacophore resulted in 3490 compounds, compared with 1162 when the merged multipharm pharmacophore was first used. When the 3490 compounds from this search were screened with the merged multipharm pharmacophore the resulting hitlist totalled 984 compounds, most likely the same compounds derived from using the reverse order of merged pharmacophores to screen the original database. However, this does not necessarily signify that the merged singlepharm pharmacophore is a less specific hypothesis than the merged multipharm pharmacophore. Rather it means that for this particular search algorithm, hypotheses containing more features are considered to be more specific. Searching the full database with each of the original hypotheses that formed the merged pharmacophores resulted in hitlists in excess of a predetermined maximum of 10 000 compounds. These 984 compounds can now be further investigated by screening with the novel pharmacophores to be developed in this study (see Sections 4.3 and 4.4).

4.3 Structure-based superligand

The superligand was generated using structure-based queries with the Unity 3D®[280] program from Tripos Inc. SiteID®[308] and Unity 3D®[280] software from Tripos Inc. was used for a one month trial period to explore some of its capabilities and their potential use in the proposed integrated drug design strategy of this research. What is presented here is the preliminary methodology for the generation of the superligand component of the integrated drug design strategy of Figure 4.1. Considerable further refinement, validation and supporting literature will be required to fully develop this computational tool.

4.3.1 Binding pocket definition

To rationally and impartially define the residues of the binding pocket to be used for the superligand generation, the SiteID module of Sybyl[309] was used to identify the residues

that form the pocket. To do this, the protein is solvated with a single layer of water molecules, as illustrated in Figure 4.12.[310]

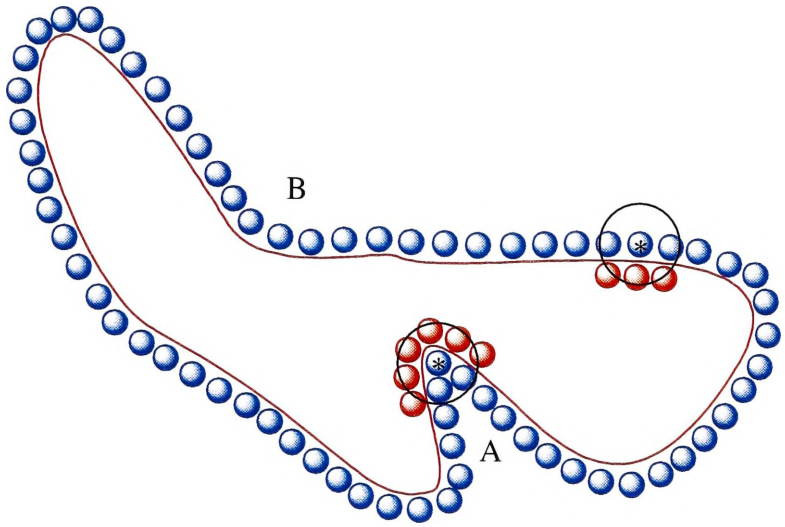


Figure 4.12 Solvation of the protein (outlined in red and represented in red spheres) with a single layer of water (blue spheres) in SiteID to determine the binding pockets, and to identify the residues of the pocket the water spheres come into contact with. The default for the radius of the black circle is 8.0 Å. Site A is identified as a pocket, and potential binding site, while Site B is only a depression and not classified as a pocket

The number of non-hydrogen atoms lying within a user-specified radius of each solvent molecule is counted. If this number exceeds a user-specified limit, the solvent atom is retained, or else it is removed. The solvent molecules that remain are those that are typically found in pockets. As seen in Figure 4.12, the first water molecule marked with a star, on the outer surface of the protein, has only three depicted red protein heavy atoms within a radius circumscribed by the black circle. The second water molecule, within an actual pocket, has six protein atoms within its radius, and will be retained. The default settings are 8.0 Å and 75 heavy atoms for the radius and atom count, respectively.

Two other significant variables in the determination of potential pockets are the *Minimum Pocket Concavity* and the *van der Waals bump scaling* factor for solvent atoms. The *Minimum Pocket Concavity*, with a default value of 2 from a range of 0 (flat) to 10 (very narrow well), determines the type of surface depressions to be considered as pockets.[310] A low value will consider nearly all surface depressions such as Site B in Figure 4.12, while a higher value will selective identify only Site A. The *van der Waals bump scaling* factor specifies the sum of scaled van der Waals radii around the solvent atom in which any atom lying within this distance will cause the

solvent atom to be excluded.[310] The default value is 0.8 Å; a higher value will mean more solvent atoms are excluded, and not considered in the calculation to identify potential pockets. Finally, all protein atoms lying within a 3.0 Å radius of each solvent atom in the identified pocket are identified as solvent accessible protein atoms of that pocket, and are considered as belonging to the active site.[310] These can be listed to define the pocket.

By virtue of the size of reverse transcriptase-inhibitor crystal complexes, it would be unreasonable and unnecessary to solvate the entire enzyme, especially considering that the purpose of this study is not to find the pocket, but rather just to characterise it. The protein was initially truncated using only residues within a 25 Å radius of the inhibitors. In the SiteID module, the truncated protein was solvated and the solvent molecules in any pockets in the enzyme were located. Two series of parameters were used to try to fully characterise the pocket. The first series of parameters used was the default, where the *Minimum Pocket Concavity* was set at 2 and the *van der Waals bump scaling* factor for solvent atoms was set at 0.80 Å, and could be considered the more strict set of parameters. In several instances the pocket was identified as being composed of two smaller pockets; the volume between them not identified as a solvent volume with the standard parameters, and these two smaller pockets were combined. In the second series of parameters the concavity was adjusted to 10 and the bump scaling factor to 0.70 Å, to try and reveal all the small recesses of the pocket not identified by the standard parameters. The second series of parameters, considered the more lenient set of parameters, may be seen as identifying and defining a larger pocket.

The residues that came into contact with the solvent spheres identified by the two series of parameters, for each crystal structure, were listed and tallied individually in Table 4.5. Each residue, for each parameter set, had a possible maximum tally of nineteen, in which case that residue came into contact with a solvent sphere in all 19 crystal structures. In the second, more lenient, series of parameters, a residue was identified as defining the pocket if its tally was greater than or equal to seven. Residues with a tally of less than seven were included if their tally in the first series was greater than seven; this occurred only with Ile142 in p51. All other residues were removed in a second truncation step to leave the sixty three residues forming the final binding pocket. This methodology ensured that a complete identification of the residues of the pocket could be made without either missing any that might be considered potentially important, or

including too many to handle easily. Six additional residues cited in literature as having contact with inhibitors, but not identified by SiteID, were included in the pocket definition. Thirty of the sixty-three residues had tallies of greater than or equal to seven in both series of parameters.

The residues of the newly defined pocket in Table 4.5 were divided into two sets, Primary Features and Secondary Features. Residues that had a tally of greater than ten as identified by the first series of standard parameters, and residues cited in the literature as having H-bonding interactions with inhibitors, were named as Primary Feature residues and the remaining residues of the pocket were designated Secondary Feature residues.

Residue details			Parameters		Lit. cited	Features	
			Series 1	Series 2		Primary	Secondary
Trp	88	A	4	11			x
Glu	89	A	8	9			x
Val	90	A	7	12			x
Gln	91	A	13	17		x	
Leu	92	A	9	13			x
Gly	93	A	12	18		x	
Ile	94	A	9	18			x
Pro	95	A	12	19		x	
His	96	A	4	17			x
Pro	97	A	1	13			x
Gly	99	A	5	15			x
Leu	100	A	13	19		x	
Lys	101	A	14	18		x	
Lys	102	A	1	1	literature	x	
Lys	103	A	9	18			x
Val	106	A	8	18			x
Gln	161	A	14	19		x	
Ser	162	A	5	8			x
Met	164	A	1	8			x
Thr	165	A	9	17			x
Leu	168	A	1	7			x
Arg	172	A	8	12			x
Val	179	A	17	19		x	
Ile	180	A	14	19		x	
Tyr	181	A	16	18		x	
Gln	182	A	13	19		x	
Tyr	183	A	7	19			x
Met	184	A	6	16			x
Asp	186	A	1	10			x
Leu	187	A	0	8			x
Tyr	188	A	13	18		x	

Val	189	A	1	6	literature	x	
Gly	190	A	3	9			x
Glu	224	A	-	-	literature	x	
Pro	225	A	-	2	literature	x	
Pro	226	A	-	-	literature	x	
Phe	227	A	7	19			x
Trp	229	A	13	19		x	
Met	230	A	6	13			x
Tyr	232	A	2	9			x
Leu	234	A	1	18			x
His	235	A	2	9			x
Pro	236	A	1	6	literature	x	
Tyr	318	A	7	14			x
Ile	380	A	3	7			x
Val	381	A	6	16			x
Ile	382	A	2	16			x
Pro	25	B	1	9			x
Leu	26	B	2	10			x
Ile	31	B	5	10			x
Thr	131	B	1	8			x
Ile	132	B	3	7			x
Pro	133	B	6	15			x
Ser	134	B	13	18			
Ile	135	B	9	17			x
Asn	136	B	4	17			x
Asn	137	B	16	19		x	
Glu	138	B	19	19		x	
Thr	139	B	19	18		x	
Pro	140	B	15	19		x	
Gly	141	B	13	16		x	
Ile	142	B	8	3			x
Arg	143	B	5	8			x

Table 4.5 List of all the residues defined as forming the binding pocket in one or more of the nineteen RT-inhibitor crystal structures, by being in contact distance (3 Å) of a solvent sphere; A refers to the p66 subunit and B refers to the p51 subunit of RT. The number of times that residue is listed as part of the binding pocket for the first or second series of parameters (see text) is tallied out of nineteen. The residues included in the pocket definition due to references in the literature are noted. All of the residues of the pocket are segregated into a primary or secondary features set (see text).

4.3.2 Consideration of protein flexibility

The need to account for the dynamic nature of a drug receptor has long been recognised as a complicating factor in computational drug design, and improvements in methodology to simulate and manipulate protein flexibility and its influence on ligand recognition have been recently reviewed.[311] A protein exists in multiple conformations, and as such a ligand may bind preferentially to any available conformation. The use of

a single protein structure can only identify ligands for that particular protein conformation.[311]

Two noteworthy methods for the consideration of flexibility were those of Sudbeck *et al.*[312] and Carlson *et al.*[311,313,314] Sudbeck *et al.* overlaid the structures of nine crystal inhibitor complexes of HIV-1 RT using the C_α of residues 97-213, and used the positions of the nine inhibitors in the overlay to create a composite binding site; a representation of the surface area encompassed by those nine inhibitors in the NNIBP.[312] This was qualitatively compared to small molecules docked and minimised in a single structure of RT. Carlson *et al.* used molecular dynamic simulations to generate a subensemble of 11 conformations, or snapshots, of uncomplexed HIV-1 integrase.[313,315] The multi-unit search for interacting conformers method (MUSIC) is a Monte Carlo simulation that was used to simultaneously calculate multiple gas-phase minimisations for hundreds of probe molecules within the active site of the 11 conformations. The binding sites for functional groups were identified by different molecular probes; methanol molecules identified hydrogen bond donor sites, and acetone probes identified hydrogen bond acceptor groups.[313] An overlay of those conformations of the active site of the protein revealed complementary binding regions that were conserved over many protein structures despite the inherent flexibility of the active site.[311] These conserved complementary regions represent a 'dynamic' pharmacophore model that allows a measure of the flexibility of the protein to be incorporated into the drug design process.[311]

In this research, protein flexibility was considered by using modified structures of the binding pocket defined previously from 19 different enzyme-inhibitor complexes. Taking the average coordinate of each atom of each residue created a modified structure for the nineteen structures, used to form an 'average' pocket structure. 'Largest' and 'smallest' pocket structures were created by using the coordinates of residues most distant from, or closest to, the center of the bound inhibitors. These pockets were not meant to represent actual conformations that the pocket might adopt, but rather were for the study of how the protein was induced to change at different locations by different known inhibitors previously crystallised. This would not only be a characterisation of the pocket flexibility, but could also be used to design new ligands that could induce multiple changes in the pocket and bind more tightly by forming more intermolecular interactions with protein residues. A 'weighted' pocket structure was also created by

taking into consideration the activity of the inhibitor that had been bound in that RT-inhibitor crystal structure. This structure would allow an indirect comparison with the techniques and results from the ligand-based pharmacophore derived from ligand structure-activity data.

To generate the coordinates for the modified pockets, the original crystal coordinates were superimposed upon the unliganded crystal structure (1dlo) in the same fashion as for the study of the conformational studies of reverse transcriptase (Chapter 3). The coordinates of all residues not included in the definition of the pocket (Section 4.31) were deleted. Taking the average value for the corresponding x, y, and z coordinates, and the temperature factor for each atom from the nineteen crystal structures generated the 'average' pocket.

In order to take into consideration the flexibility of reverse transcriptase without resorting to expensive dynamics calculations, the 'largest' and 'smallest' pockets were generated by considering each individual residue defining the pocket, and how far it 'expanded' to create the pocket. The x, y and z coordinates of each atom of all the inhibitors of the crystal structures were averaged to generate a central reference point (x_1,y_1,z_1). This method is reasonable if one bears in mind that these inhibitors were also overlaid by the initial superimposition of all the original crystal coordinates; this overlay can be seen in Section 3.7.1 Inhibitor binding, Figure 3.49.

The atoms of the residues defining the pocket were segregated into two groups, backbone and sidechain atoms, for each residue. These were averaged (x_2, y_2, z_2) so that each residue had one point which represented its backbone and one which represented its sidechain. The distance between the central inhibitor reference point, and the backbone and sidechain point was calculated.

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

The crystal structure with the minimum and maximum distance values for the backbone and sidechain of each residue was noted, and the complete coordinates for this residue from this crystal structure were then used for the modified pockets (for further details see Chapter 6 Experimental). Thus four new pockets were generated; the smallest and largest pockets based on how far the backbone was away from the central inhibitor

reference point, and the smallest and largest pocket based on the sidechain displacement from the reference point.

The activity 'weighted' pocket was generated to directly parallel the ligand-based pharmacophore generated in Catalyst (Section 4.2). The inverse logarithm of the activities of the crystal structures used in the generation of the ligand-based pharmacophores was normalised by giving 1rti, the least active inhibitor, a weight of 1.00 as in Table 4.6. The *x*, *y*, *z* coordinates of each atom of each crystal structure was multiplied by that crystal structure's weighting. The sum of these coordinates was calculated and then divided by the sum of the weightings used in that calculation, generally 68.2851. If the sidechain of a residue was missing from one particular crystal structure, then it was not included in the calculations, and the sum of the coordinates was divided by a reduced total weighting.

It had been noted in the conformational study of the binding pocket (Section 3.7) that the sidechain of Tyr181 for all the crystal structures, excluding rti and rt3, were clustered together. In rti and rt3, the tyrosine had a 90-100° swivel about the C_β atom. The 'average' pocket of all the crystal structures excluding 1rti and 1rt3 was generated, as well as the 'average' pocket of just rti and rt3. It was noted, retrospectively, that although rti was included in the generation of the 'weighted' pocket, it was not expected to play a dominant role in the weighting process, having the lowest activity and a weighting of 1.00. The ligand rt3 was not included in the 'weighted' pocket as it had previously not been used for the ligand-based drug design (see Section 4.2.1 Training set selection).

The results of these binding pocket modifications resulted in several interesting 'binding sites' for further study. In particular, the comparison of the superligands derived from the four 'flexible' pockets; the smallest and largest backbone-based pockets and the smallest and largest sidechain-based pockets, were certain to reveal more information about how different modelled conformations of the pocket can affect the binding of ligands. However, the brevity of the trial period for the Tripos software curbed extensive study of all the proposed modified pockets, and further research towards the generation of the superligand was undertaken only with the 'weighted' modified pocket.

	IC_{50} (nM)	log (IC_{50})	1 / log (IC_{50})	Weighting
bqm	80	1.90	0.53	2.5325
hnv	4.3	0.63	1.58	7.6082
klm	1100	3.04	0.33	1.5847
rev	1.5	0.18	5.68	27.3696
rt1	12	1.08	0.93	4.4659
rt2	6	0.78	1.29	6.1936
rt5	200	2.30	0.43	2.0945
rt6	150	2.18	0.46	2.2148
rt7	280	2.45	0.41	1.9694
rth	400	2.60	0.38	1.8522
rti	66000	4.82	0.21	1.0000
vrt	84	1.92	0.52	2.5046
vru	5	0.70	1.43	6.8952
			Total	68.2851

Table 4.6 Calculation of the crystal structure weightings.

4.3.3 Superligand generation with Unity queries

Unity 3D®[280] is a rapid database searching program for the identification of structures matching ligand-based pharmacophores and receptor-site constraints. Its recent emergence as a computational tool has meant that published literature of the use of the program has been exceedingly sparse. Only one reference has been published with its ligand-based pharmacophore use,[316] and one recent reference to its use with receptor-site constraints.[317] The basic premise of its structure-based functions is illustrated in Figure 4.13, where the inverse pharmacophoric feature of a chemical function of a particular residue of the receptor or binding pocket is placed at an optimal position for a ligand functional group to interact with that residue. Thus the carbonyl of a residue in the receptor site, which may possibly function as a hydrogen bond acceptor, has a hydrogen bond donor pharmacophoric feature (purple sphere) matched to it within the binding pocket. The collection of pharmacophoric spheres is referred to as a structural query.

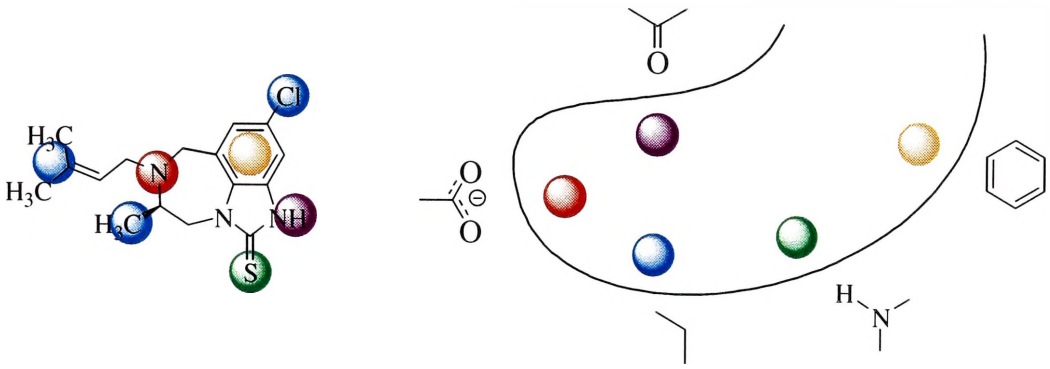


Figure 4.13 (Left) Coloured spheres represent the pharmacophoric features of a chemical structure. (Right) Coloured spheres represent the inverse pharmacophoric features of a protein structure. Beige spheres are ring aromatic features, green spheres represent hydrogen bond acceptor features, purple spheres represent hydrogen bond donor features, red spheres represent positive ionisable functions, and blue spheres represent hydrophobic aliphatic functions.

Additionally Unity 3D allows the user to build structural queries from molecules, molecular fragments, and pharmacophore models. Multiple excluded volumes representing the binding site can be used to exclude structures that may have steric problems binding within the receptor, or help define the containing volume of the receptor cavity.[318] The flexible three-dimensional database searching function of Unity 3D allows these structural queries to be used to search databases to find molecules that can achieve a matching conformation to the query, regardless of the stored conformation in the database.[318] Significantly, Unity 3D also allows partial match directives, which tolerate searches for ligands containing only some of the features specified in the query. The generation of the structural query may be best summarised in Figure 4.14, actually derived from a description of the Structure-Based Focusing module of Cerius[2®281.282] The three-dimensional database in which the query searches may be user-created, public (National Cancer Institute), or commercially available (Maybridge, Derwent's World Drug Index, MDL Drug Data Report, and Comprehensive Medicinal Chemistry databases). Thus these latest additions to the range of computational tools available are ideal substitutes for the method of superligand generation previously proposed by the integrated drug design scheme of Figure 4.1. In our case, the generation of the Ludi interaction map and clustering of the interaction site vectors are exactly the same steps proposed to generate the superligand in Figure 4.1.

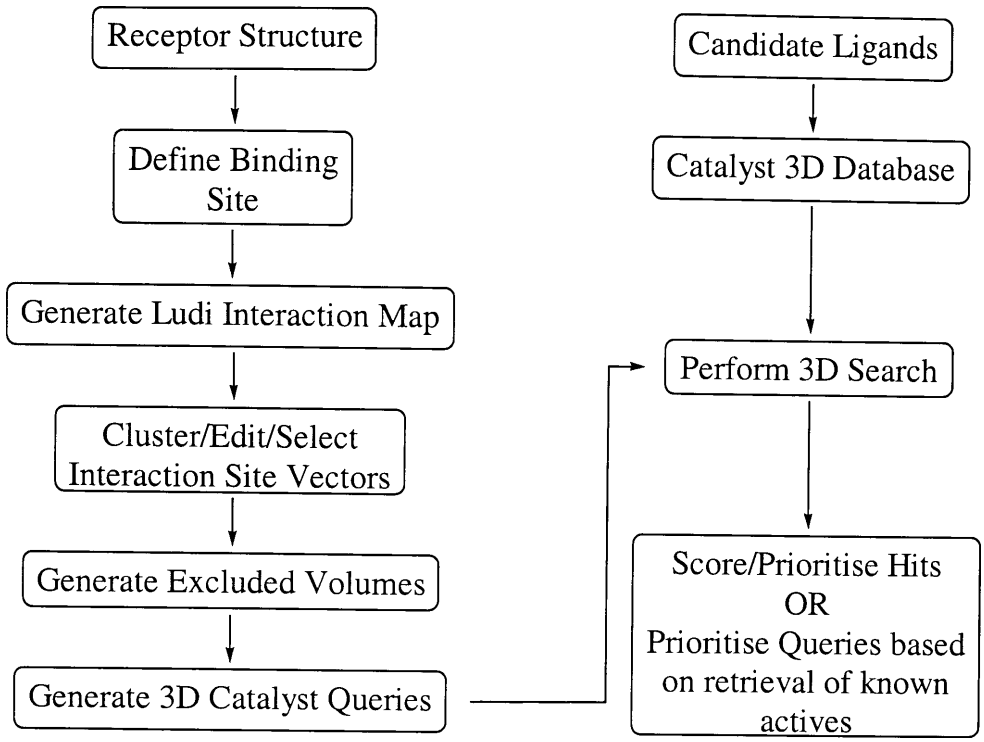


Figure 4.14 Basic schematic of the process to generate and use structural queries.[282]

Unity can create numerous different types of features to generate a structural query. These can include atoms, bonds, lines, planes, centroids, extension points, hydrogen bond sites, hydrophobic sites, and spatial features which define spheres, lines, planes, caps and toruses in which a feature must lie for it to be considered a 'hit'.[318] However, the most common features used in automated feature searching of receptors and active sites are the hydrogen bond donating and accepting, and hydrophobic features, which are illustrated in Figure 4.15. The residues of the 'weighted' binding pocket were searched for all possible hydrogen bond donor and acceptor features, and these were displayed graphically as extensions from the protein residue. A manual selection process was then required to choose those features suitable for inclusion in the structural query. Donor and acceptor features projecting into the binding pocket were selected, while those that projected out of the pocket, and most likely into a space of the protein that had been cut away during the initial preparation and definition of the pocket (Section 4.3.1), were deleted. Hydrophobic residues could not be selected automatically, and so all of the aromatic residues listed in the pocket definition (Table 4.5) were selected to generate the corresponding hydrophobic feature of the ligand in the binding pocket. Aliphatic amino acid residues were not selected, as this would have increased the already extremely large number of features to an ungainly size, and hydrophobic aromatic features in the binding pocket had already been noted as being of particular importance.[30,215]

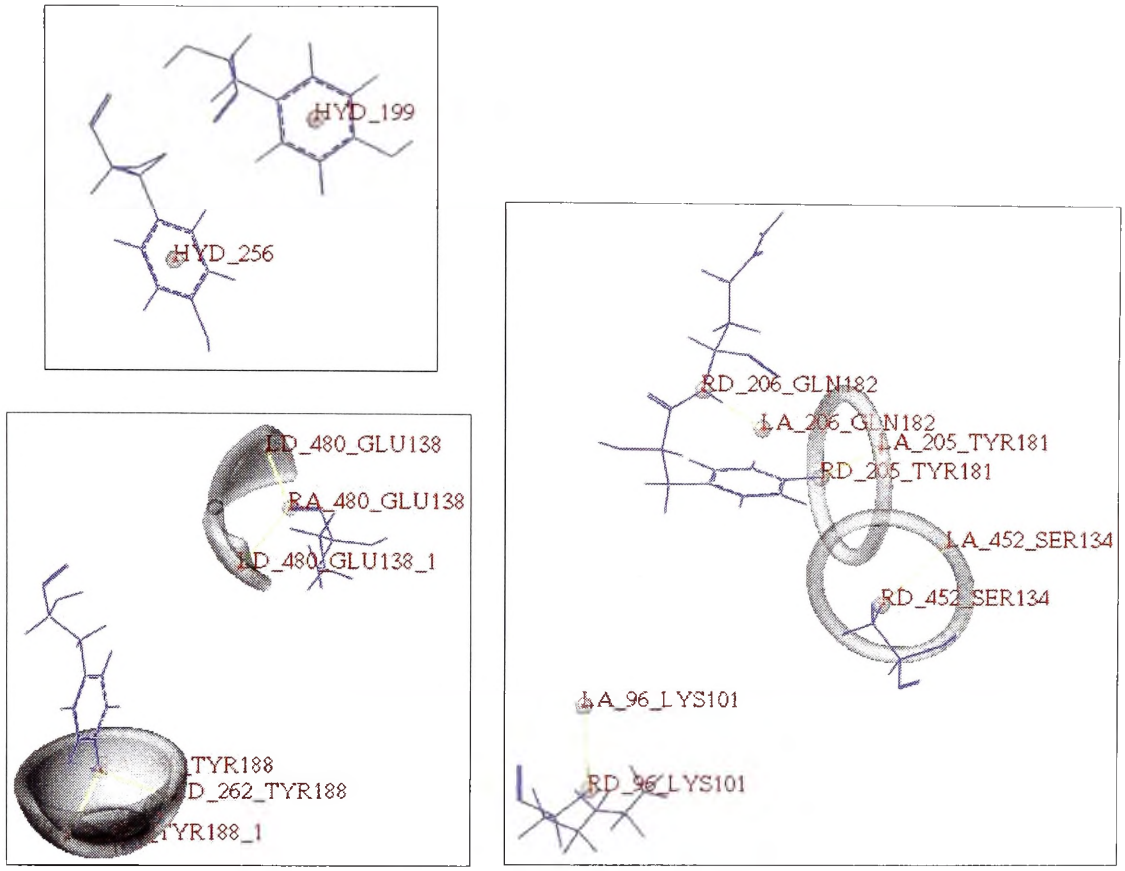


Figure 4.15 (Left, top) Hydrophobic features for Tyr181 and Tyr188; the residues of the protein are shown in blue and the features of the query are shown in grey. (Left, bottom) Hydrogen bond acceptor functions on the residues of the protein project the possible spatial volume (bifurcated cap for Glu138 and cap for Tyr188) in which a hydrogen bond donor function on the ligand could lie to make an intermolecular contact with the protein. (Right) Hydrogen bond donor functions on the residues of the protein project the possible spatial volume (torus for Ser134 and Tyr181, and small spheres for Lys101 and Gln182) in which a hydrogen bond acceptor function on the ligand could lie to make an intermolecular contact with the protein.

Another significant element of a structural query is the inclusion of excluded volumes. These are generally placed to represent residues of the binding pocket to ensure that, during database searching, ligands are not chosen that will fit all features but intrude into a space occupied by the receptor, and thus be unlikely to bind due to steric problems. A well defined excluded volume for a receptor will also function well as an inclusive volume, which forces a ligand to be within that volume to register as a hit, without being as restrictive. A useful parameter of the excluded volumes feature is that of the van der Waals radius that allows a user to specify the radius of the excluded volume sphere around each atom. By reducing this value, one in fact allows for protein flexibility to some degree.

Three queries were created from the 'weighted' pocket to use to sequentially screen the NCI database as implemented in Unity 3D®, version 4.1.[280] Ideally, one would generate

a single, all-encompassing, structural query, equivalent to the superligand concept, with all the pharmacophoric features of the primary and secondary features sets of Table 4.5 to be used for database searching. However, this becomes computationally infeasible, as there are too many conformational and binding possibilities to explore for all save very small databases. This is especially the case when one considers the best and preferable form of searching also considers the flexibility of the ligand, searching for an appropriate conformation to fit the pharmacophoric query. In addition, one of the emerging and promising functions of database search algorithms is the capacity for partial-match searches, where a minimum and maximum number of features that are required to map to the pharmacophore, can be set.[280,302].

The first query generated from the 'weighted' binding pocket used the spatial hydrophobic features of Tyr181 and Tyr188, and the excluded volumes of all residues in the secondary features set with the default van der Waals scaling factor of 1.0. The two tyrosines act as anchor or reference points for the database ligand searching to narrow the possibilities, while the excluded volumes limit the number of very large hydrophobic-feature containing ligands that would otherwise be considered as hits. Residues 181 and 188 were used as anchor points, as the literature has shown that these two residues provide the most significant intermolecular binding interaction with the ligand in the pocket.[30,215] This boxing-in effect of the excluded volumes is illustrated in Figure 4.16 (left). This is not an ideal search method as the residues of the secondary features set have been identified in some RT-inhibitor complexes as being solvent accessible, and hence having possible interactions with the ligand, or potential ligands, in the binding pocket (see Section 4.3.1 and Table 4.5). However, this compromise was necessary in consideration of computational time and power.

The second query was the same as the initial query with the addition of all residues of the primary features set, excluding Tyr181 and Tyr188, defined as excluded volumes with a van der Waals scaling factor of 0.5. This query further increased the stringent criteria required of ligands if they were to be considered as potentially capable of binding into the pocket, and is illustrated in Figure 4.16 (right). The concession of using a van der Waals scaling factor of 0.5 allows potential ligands to approach closer to residues of the primary features set. These were residues that were identified as being in contact with a solvent sphere in at least ten of the RT-inhibitor-complexes, or noted

in the literature as being significant, and are more likely to be involved in intermolecular interactions with the ligand.

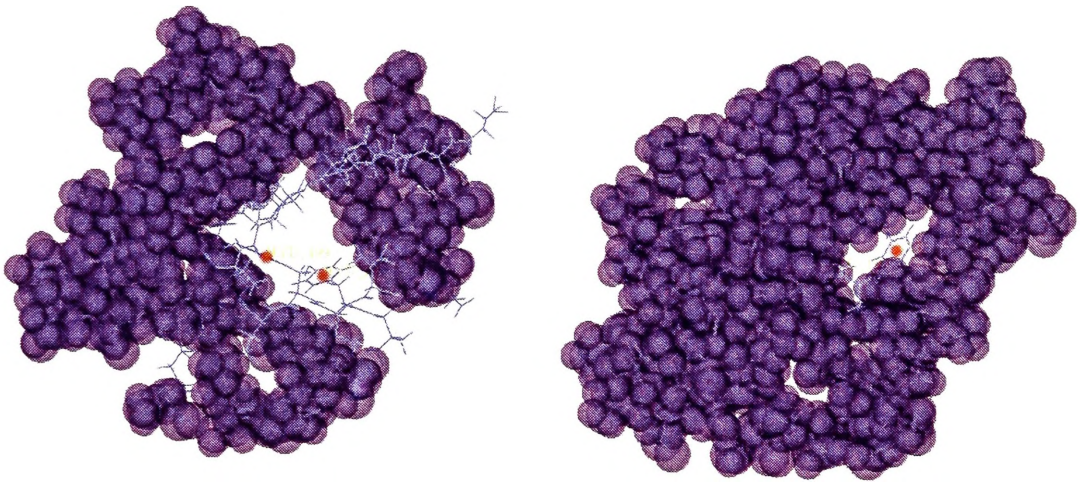


Figure 4.16 (Left) Structural query 1 with the residues of the secondary features set shown as purple excluded volume spheres, van der Waals radius 1.0 Å. The two hydrophobic features of Tyr181 and Tyr188 have been annotated as red spheres. (Right) Structural query 2 with the residues of the secondary features set shown as purple excluded volume spheres, van der Waals radius 1.0 Å, and the residues of the primary features set also shown as purple excluded volume spheres, van der Waals radius 0.5 Å.

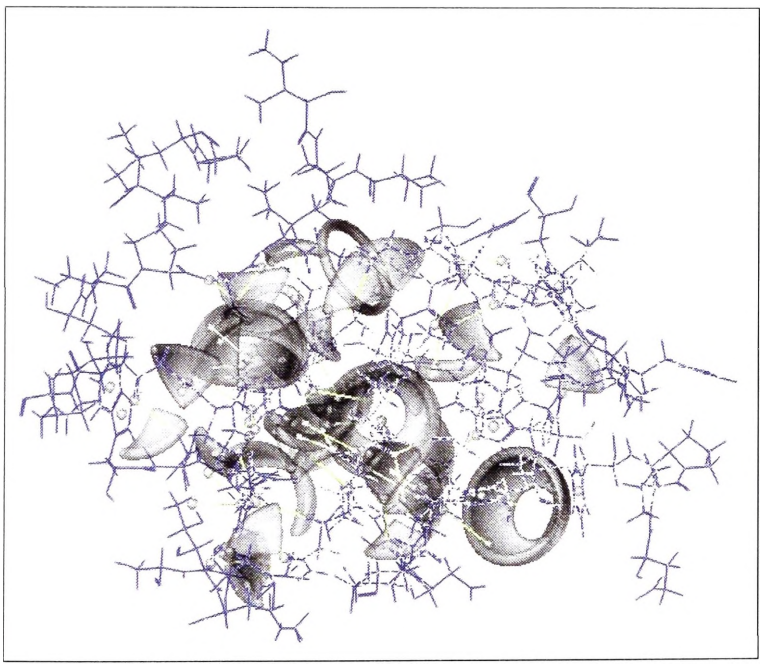


Figure 4.17 Structural query 3, the superligand of the 'weighted' binding pocket. The residues of the defined binding pocket are seen in blue, while the features of the structural query are seen as grey volumes.

The third query, illustrated in Figure 4.17, was composed of all features generated, including 52 donor, 46 acceptor and 10 hydrophobic features from the appropriate residues from both the primary and secondary features sets. In fact, this third query is the superligand initially proposed as the structure-based pharmacophore, as it describes with pharmacophoric features, all of the potential binding interactions the residues of the pocket are capable of.

4.3.4 Database searching with Unity queries

All three queries were used to search an NCI database using Unity's Flex Query option, which allows database ligands to be considered flexible while searching for matches to the query. The results from the first query were saved as a separate database, and the second query was then used to search this subset database. The results from the second query search were also saved as a subset database for the third query to search. The third query was performed using Unity's Partial Match Constraint Dialog, which allows a search to be conducted where a match will contain only some of the features specified in the query. The results of the database searching are presented in Table 4.8.

The first structural query was performed on an NCI database of 117 649 compounds. The requirement of the minimum number of features of two ensured that all ligands had two hydrophobic features spaced apart approximately the same distance as the two tyrosines of the 'weighted' pocket. The macro screen removes from consideration any compounds in the database that did not meet this first criterion of having two hydrophobic functionalities. Of the remaining 71 951 compounds, 27 329 were considered as hits; mapping both hydrophobic features and being within the binding pocket circumscribed by the excluded volumes of the secondary features set. A number of the searched compounds (682) were noted as timed out; these compounds were likely to have contained numerous flexible bonds that needed extensive conformational searching to find a fit to the structural query. A fit was not found for these compounds after a preset time of 90 seconds, and the search algorithm moved on to the next structure. As these compounds were likely to contain too many rotatable bonds to be of interest, they were discarded and not added to the subsequent search volume.

The second structural query further increased the strictness of the criterion for ligand binding in the pocket, and reduced the number of potential compounds to 11 110. The number of compounds that timed out as the search algorithm investigated conformations so the compound that could fit into the reduced binding pocket was 701. Once again these were excluded from the next collection of compounds to be screened. At this point, one could say with some degree of confidence that all 11 110 resulting compounds could potentially bind in the pocket as they all had two hydrophobic features capable of possibly interaction with Tyr181 and Tyr188, and were all of approximately suitable size.

	Query 1	Query 2	Query 3
Features	Tyr181 and Tyr188	Tyr181 and Tyr188	Primary Features Secondary Features
Excluded Volumes	Secondary Features (vdW 1.0)	Primary Features (vdW 0.5) Secondary Features (vdW 1.0)	not used
min. features	2	2	8
max. features	15	15	16
database size	117 649	27 329	11 110
macro screen	71 951	27 325	11 110
hits	27 329	11 110	172
timed out	682	701	7537

Table 4.8 Results of consecutive searches with different structural queries to screen the NCI database. The Features row denotes features in the query that were available for a ligand to fit or map to. The Excluded Volumes describes the residues that were set as excluded volumes, and the van der Waals scaling factor of those residues. The min. and max. features are the values used in the partial match search algorithm to specify the minimum and maximum features, respectively, that a ligand needs to fit to be considered a hit. The remaining numbers are the results of the database searching.

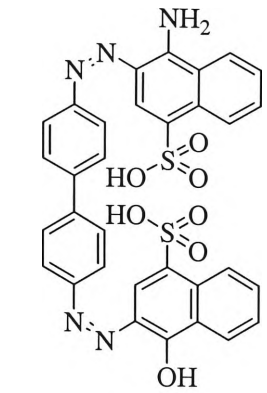
The final structural query endeavored to reduce this hit list down to a reasonable size of less than two hundred, for further manual selection and consideration. In this search, all of the features identified as having possible binding interactions with the residues of the primary and secondary features sets were considered to form part of the structural query. The extremely high number of compounds that timed out, 7 537 compounds, was due to the large number of features in the query. Imposing a minimum of eight features that needed to be mapped onto the structural query meant that the search algorithm needed to search multiple conformations of the compound to see if they could fit onto multiple possible combinations of eight features. This was a considerable computational undertaking, and explained the high number of compounds that were not able to find a fit in the allocated time span of 90 seconds. Ideally, these timed-out compounds should be searched with longer allocation times per structure. Conversely, the same previous argument could be made that these timed-out compounds were likely to be too flexible for consideration. Regardless, the number of returned hits was now at a reasonable size and no further database searching was conducted.

4.3.5 Compound selection for biological testing

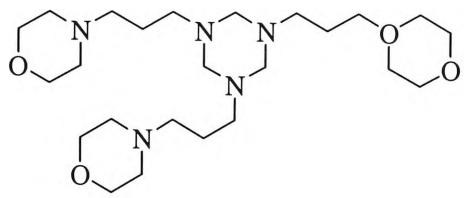
The activities of the 172 compounds, that were identified as potential ligands for the 'weighted' pocket, were estimated by importing them into a spreadsheet in Catalyst, and using the Score Hypothesis tool. Two ligand-based pharmacophores generated in Catalyst were selected to estimate the activities of the compounds of the hitlist with the Score Hypothesis tool. The highest ranking hypotheses from the singlepharm pharmacophore, where the *Minimum Total Features* parameter was set at 1, and the multipharm pharmacophore, where that parameter was set at 4 (see Section 4.2.4 Hypothesis analysis and selection), were selected. These estimated values were only one small point of consideration in the selection of compounds to be requested from the NCI for antiviral and anti-malarial testing. Additional criteria included the structural diversity of ligands, both rigidity and flexibility in the ligand, size and molecular weight. The inclusion of familiar structural motifs encountered in antiviral natural products and current anti-HIV drugs, such as multiple aromatic features, a possible 'butterfly' conformation of the compound, and dimeric structures were also taken into account. With these considerations in mind, ten compounds were requested from the NCI, of which seven were donated. These seven compounds, shown in Figure 4.18, were sent for both anti-HIV and antimalarial testing (see Section 2.9).

As the database searched was that of the NCI, its Development Therapeutics Program (DTP)[319] had available the anticancer and antiviral screen results of some of the compounds of that database. A search of the structures of Figure 4.18 showed that compounds 7231 and 55233 had IC_{50} values of $> 1.00 \times 10^{-4}$ and 3.17×10^{-4}, respectively. The activity of 7231 was noted as active, while that of 55233 was inactive. No data was available for the remaining compounds. Notably, NCI anti-HIV screens use whole cell assays. These results provide a preliminary and tentative validation of the structure-based query.

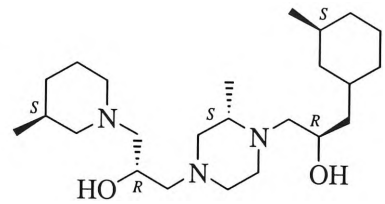
The DTP searchable database[319] was only used after the selection of the compounds for biological testing, however, it remains a useful tool in the assessment of larger hitlists obtained from database searching. Compound selection should still be based on some of the criteria previously noted, and on fitting to select pharmacophores.



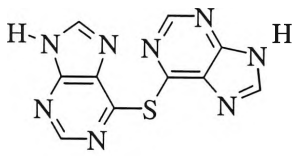
7231
2429-70-1



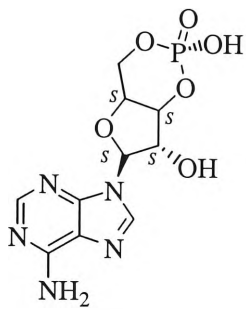
15743
5434-22-0



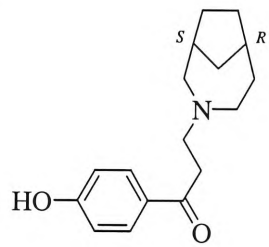
55233
6624-14-2



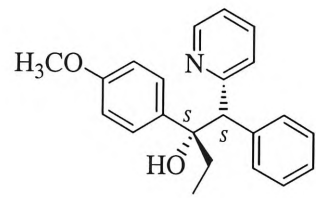
39055
90947-51-6



143670
51867-88-0



95551
73826-34-3



24238
16164-60-6

Figure 4.18 Structures of NCI donated compounds selected from database search with NSC code number and CAS registration number.

4.3.6 Structural queries in Catalyst

In addition to the CatHypo tool for automatic hypothesis generation, Catalyst® also has a utility for the generation of 'manually' constructed hypotheses, called Hypoedit. The Cartesian coordinates and constraints of all typical pharmacophore features can be described, as can excluded volumes. Constraints describe the volume and locations of the spheres typically seen with the Catalyst pharmacophores. This utility has been previously used to incorporate structural x-ray crystallographic data in the form of excluded volumes to ligand, or docking-derived pharmacophores.[284,285,289] The default settings specify that features have a constraint tolerance of 150 picometres (pm) on the ligand, 200 pm for the corresponding interacting feature of the receptor, and 250 pm for excluded volumes.

Using this utility, one could reconstruct the structural queries from Unity 3D in Catalyst and not only compare this structure-based pharmacophore to the ligand-based pharmacophores previously generated in Catalyst (Section 4.2), but also fit ligands to the structural query. Given that the structural queries from Unity only included

hydrogen bond donor, acceptor, and hydrophobic functions and excluded volumes, these were the only types of features used.

The first step in the process of translating the structural query to a Catalyst hypothesis involved retrieving the structural query from Unity 3D. The third query used in the database search was used to test the feasibility of this translation. Structural queries can be saved as text files, and provide the coordinates for each feature of the query. A donor atom on the receptor is designated RD and the corresponding location of the acceptor atom on the ligand is designated LA, coordinates are provided for both RD and LD and together they describe a hydrogen bond acceptor function for a ligand-based pharmacophore. The same is true of the hydrogen bond donor function that is described by an acceptor atom on the receptor, RA, and a corresponding donor atom on the ligand, designated LD. Hydrophobic features and excluded volumes only have one set of coordinates rather than two.

Two additional considerations had to be made in the selection of features to be 'translated' with the Hypoedit utility. One concerned the use of bifurcated features in Unity, that is, features where an acceptor or donor site on the receptor projects two separate donor or acceptor sites, respectively, to possible corresponding features on the ligand. A bifurcated feature is illustrated in Figure 4.15 (left, bottom). Catalyst was unable to deal with these bifurcated features, believing that there were too many constraints to the one feature. With these bifurcated features from Unity, only the coordinates from the first feature of the bifurcation were used. The second consideration was related to the first, and involved receptor features that could act as both hydrogen bond donors and acceptors. This occurred twice, and in each case the feature where the ligand contained the acceptor features was used preferentially to the feature where the ligand contained the donor atom. This was due to the greater number of donor features (52) compared to acceptor features (46) in the structural query in Unity, including bifurcated features.

A Hypoedit script file is provided with the utility (Appendix 7.14), and this is copied and edited with the coordinate data of the features extracted from the Unity structural query. Two reconstructions of the Unity structural query were made with the Catalyst Hypoedit utility, one with features, seen in Figure 4.19 (top) (Appendix 7.14.1), and one with both features and exclusion volumes, seen in Figure 4.19 (bottom). This latter pharmacophore has 87 excluded volumes from the excluded volumes defined by the

residues of the secondary features set (Table 4.5). A maximum limit of 1000 excluded volumes and 100 features has been noted for the Hypoedit utility.[301]

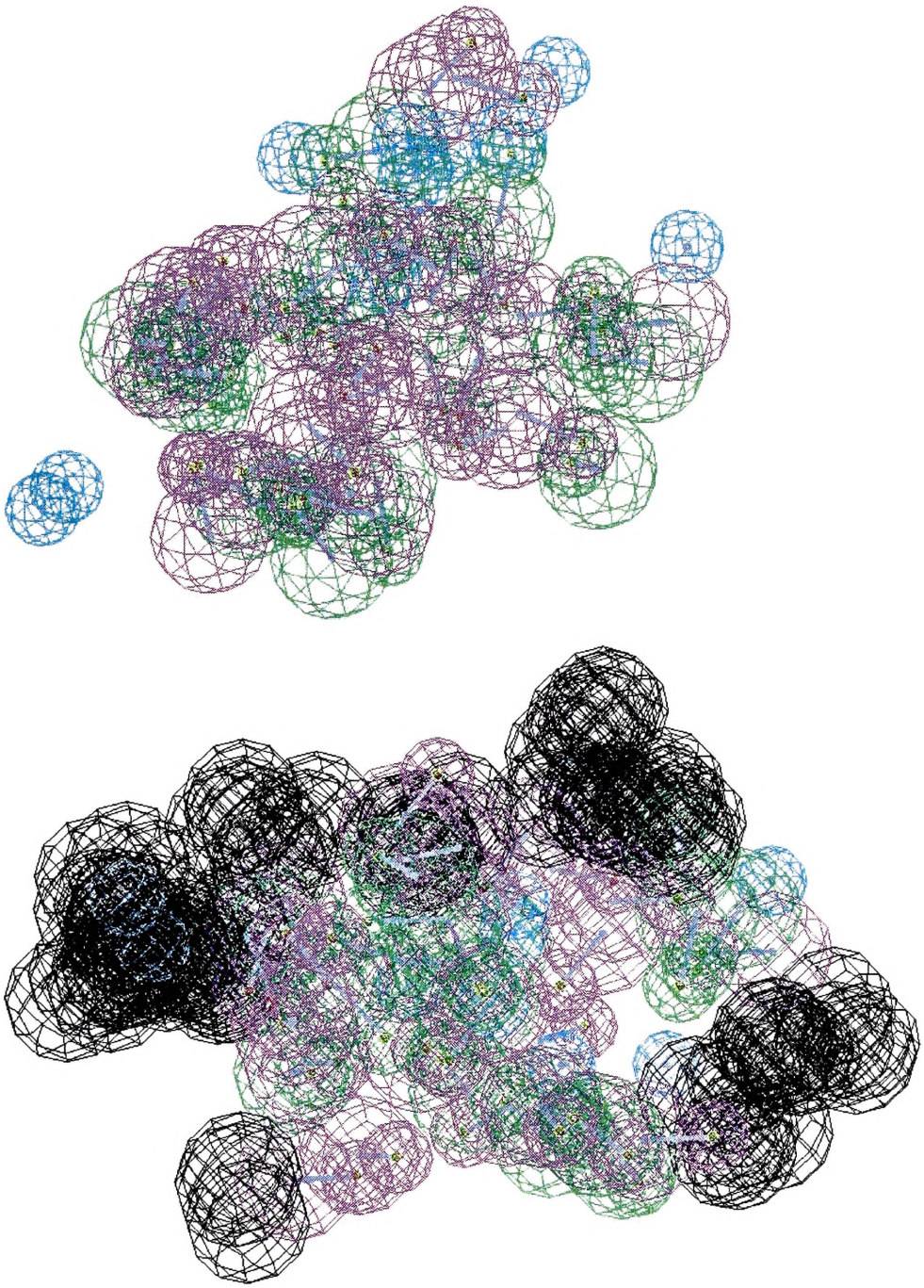


Figure 4.19 (Top) Structural query from Unity seen as a Catalyst hypothesis. Green wire-cages represent hydrogen bond acceptor features, purple represents hydrogen bond donor features and aqua represents hydrophobic features. (Bottom) The same structural query with excluded volumes. Black wire-cages represent excluded volumes.

Unfortunately, attempts to use the structural query pharmacophore generated with Hypoedit and compare the merged singlepharm and merged multipharm pharmacophores were unsuccessful as the only common feature between the structural query pharmacophore and the merged pharmacophores was the green hydrogen bond

acceptor feature. Furthermore, database searching with the structural query pharmacophore was not possible with the particular version of the software used,[307] as it did not allow the necessary partial match searching. The search algorithm of Catalyst 4.5 requires that a ligand at least have the same number and type of features as the searching pharmacophore to even be considered

4.4 Future directions for integrated drug design

This research, endeavoring to realise the optimal synergism of ligand and structure based design principles, has only taking the first step towards fully utilising all the available information for drug design. No studies towards the actual integration of the ligand and structure based pharmacophores have been attempted, and admittedly no tangible or prospective strategies to integrate the ligand and structure-based pharmacophores to create the proposed combiphore have been formulated. This remains a significant and worthy challenge for future research efforts.

Medicinal agents designed from the combiphore are proposed to be able to consider both present data about inhibitors known to be active at a binding site, as well as all potential interactions the binding site may be capable of. The complete exploration of the binding interactions of a pocket provides the combiphore drug strategy the ability to target the residues of a protein that may not previously have been utilised, and aid in the design of novel agents effective against mutant strains of proteins resistant to present drugs. Potentially, the mode in which mutations of residues in the binding site diminish or eliminate inhibitor activity may also be easily studied and quantified using the combiphore. Each feature of the combiphore can be weighted; these weightings can be increased, reduced or abolished as mutations in the pocket differentially affect binding interactions. Combiphores may be an effective tool in studying how different ligands acting at one binding site can exert both agonistic and antagonistic effects, as they can include agonistic and antagonistic pharmacophores from ligand-based design.

Naturally, these potential utilisations can only be possible after much more thorough study and research with the two foundations of the combiphore, the ligand-based pharmacophores and the structure-based queries. In the ligand-based studies, further refinement of the individual pharmacophores and database searching strategies are required to improve the quality of the hits from a database search. The inclusion of excluded volumes may be possible by docking an individual hypothesis, or a merged

hypothesis, into an actual x-ray crystal structure of the pocket. Docking using the 'weighted' pocket created for the structure-based studies would provide an interesting study. The coordinates of the residues of the pocket could be transformed relative to the pharmacophore, and added to the pharmacophore in Catalyst with the Hypoedit utility. The docking of pharmacophores into individual x-ray crystal structures may also be interesting to see how the crystallised ligand of the RT-inhibitor complex overlays with the pharmacophoric features.

In the structure-based studies, potentially valuable and interesting research should be conducted into docking studies of the ligands from the Unity database searches in the newly created and modified binding pockets. These would help in the identification of the most likely lead compounds to consider for synthesis and biological testing. Conversely, compounds identified from the ligand-based pharmacophore searches should be either docked into these structure-derived queries with various docking algorithms, or fitted with pharmacophoric search algorithms.

Pharmacophore, or structural query cross searching should also be explored further than the preliminary attempts in the selection of compounds for biological testing from the Unity 3D database searching (Section 4.3.4). That is, searching a selection of compounds considered hits on the ligand-derived pharmacophore with the structural query pharmacophore, and vice versa. Partial match querying is absolutely required for this process, as the structural query pharmacophore would be expected to have too many features.

It would not be an exaggeration to state that the inventory of research that could potentially arise from the basic concepts of the superligand and combiphore presented in Figure 4.1 is significant. In reality such ambitions have already produced such computational tools as Unity 3D®,[280] and the Structure-Based Focusing module of Cerius[2®281] The future for this research therefore remains as vast and open as its potential was seen several decades ago with the advent of computer aided ligand design.

Chapter 5

Conclusions and Future Directions

5.1 Synthetic studies

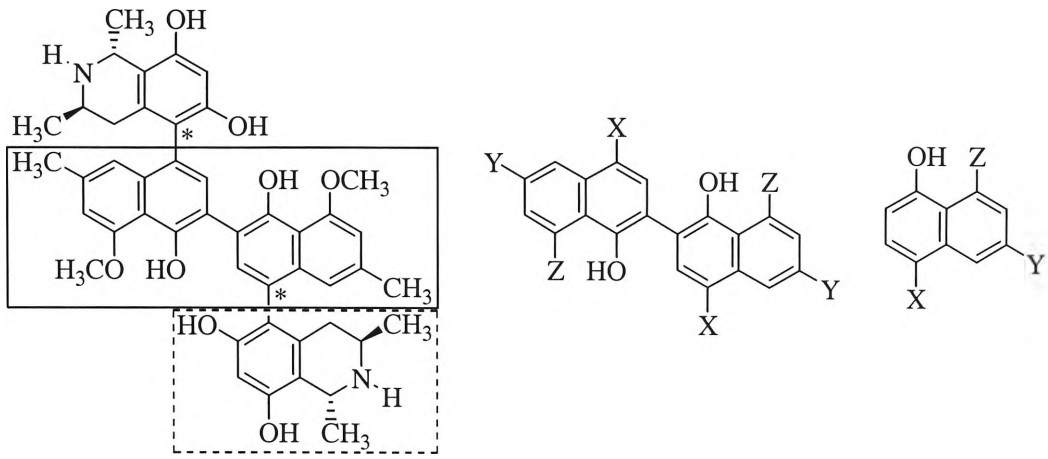
The research undertaken to synthesise analogues of the antiviral michellamine alkaloids (Figure 5.1) established synthetic methodologies for biaryl cross-coupling and homo-coupling reactions. The Suzuki reaction was used successfully to form the biaryl bond to replace the tetrahydroisoquinoline moiety of the michellamines, and to construct the central biaryl axis to yield the dimer analogues. Coupling reactions using the Stille methodology, also yielded cross-coupled and homo-coupled products, however the poor yields made this an unattractive synthesis.

Cross-coupled products in the Suzuki reaction were synthesised with the boronic acid or ester moiety on either the naphthyl ring system that would form the michellamine skeleton, or on the moiety replacing the tetrahydroisoquinoline. The most common catalyst used was dichloro[1,1'-bis(diphenylphosphino)ferrocene]palladium, $PdCl_2$(dppf), as this accelerated the coupling, and ensured reaction times were of the order of 24 hours, compared to tetrakis(triphenylphosphine) palladium(0), $Pd(PPh_3)_4$, with which most reactions required at least 36 hours to complete. Another catalyst used was palladium acetate, for the synthesis of the pyridine-substituted analogue. The most common solvent for the coupling reactions was THF, and only the synthesis of the pyridine-substituted analogue specifically required the use of DME. The choice of base was dependent on the reaction; weaker bases were generally used in the cross-coupling reactions to minimise homo-coupled by-products, and stronger bases were used in the homo-coupling reactions to accelerate the formation of the product and minimise deboronation. Deprotection of the synthesised monomeric and dimeric analogues was

performed to yield the desired potential inhibitors. However, future synthetic studies should consider optimisation as the current conditions produced unsatisfactory yields for the last reaction in a multi-step synthesis.

The synthetic studies established and, in some measure, optimised a methodology for the cross-coupling of aryl and alkyl substituents to a brominated naphthyl moiety, and the homo-coupling of that *ortho*-brominated cross-coupled product to form binaphthyl analogues of the michellamines. One of the initial aims of this research, to determine if these monomeric and dimeric analogues were active against HIV-1 reverse transcriptase or showed activity in antimalarial screens, was not wholly achieved at this time, however, biological testing results will be available soon.

The most immediate next step of this research is the completion of the synthesis of the dimeric and monomeric series of designed inhibitors; in particular the pyridine substituted compounds. The monomeric and dimeric phenyl, naphthyl, methyl, pyrimidyl and pyridyl X-substituted inhibitors of Figure 5.1 also require transformation of their Y-substituents; ethyl ester to carboxylic acid, methyl alcohol, and methyl groups. Hydroxy substitutions in this position should also be considered to introduce significant and necessary structure-activity relationship (SAR) data to add to those inhibitors currently being tested, as should varying the methoxy group to Z-substituents. The synthesis and design of further michellamine analogues and antiviral inhibitors will require the contributions from SAR data, and ligand- and structure-based drug design.



Michellamine (**26**)

Figure 5.1 The michellamine alkaloid and proposed substituent points for analogue synthesis.

5.2 Conformational changes of the reverse transcriptase enzyme

The studies of the conformational changes of the reverse transcriptase enzyme induced by substrate and inhibitor binding established an objective, and quantitative method to analyse a large number of x-ray crystallographic structures. In this research, crystal structures with and without DNA, and non-nucleoside inhibitors bound were used. In the substrate-bound crystal structures, the polymerase active site showed displacements from the position adopted in the unliganded structure that are most likely explained as movements to accommodate the double stranded nucleic acid and the incoming nucleotide, to prime the enzyme for catalysis of the chemical reaction. Similar small, but notable, conformational changes were observed in the RNase H active site and the primer and template grips.

The inhibitor-bound crystal structures displayed quite large displacements in the studied active sites and structural motifs. In the polymerase and RNase H active sites, the conformational changes relative to other structural features and motifs, as well as between the residues of the active sites, saw a change in the coordination complexes that the catalytic residues may be able to adopt. These coordination complexes are quite involved, requiring the exact positioning of the residues of the active site, nucleic acid template and primer strands, incoming nucleotides and Mg^{2+} ions for the successful chemical steps of polymerisation and hydrolysis. The primer and template grip motifs demonstrated distinct displacements from the positions adopted in the unliganded and DNA-bound structures, occupying some of the space of the double-stranded DNA in the substrate-bound structures. The most direct inference from these results was that non-

nucleoside inhibitor binding affected the intermolecular interactions between the magnesium ion(s), the incoming nucleotide, the template and primer strands of nucleic acid, and the corresponding enzyme grips. This may either make it impossible for the nucleic acid to bind and/or hinder its mechanism of translocation during the synthesis of the DNA chain.

The non-nucleoside inhibitor binding pocket was also analysed in substrate- and inhibitor-bound crystal structures and changes were found to be predominantly caused by the displacement of Tyr181 and Tyr188, and residues that the binding pocket has in common with the primer grip. The supposition from these observations is that the displacement of the residues of the pocket, in particular the swivelling of the two tyrosines, creates the space that becomes occupied by the non-nucleoside inhibitor, and in doing so shifts the position and conformation of neighbouring residues. These structural disturbances have both short- and long-range consequences, as observed in the two active sites and nucleic acid grips studied.

Consideration of the conformational changes in the substrate and inhibitor-bound structures resulted in the postulation that the mechanism of non-nucleoside inhibition involved not just one structural motif, such as the PAS, as has been suggested in the literature. More likely, the inhibition and activation of RT is achieved by integrated and synchronous conformational changes of all significant structural motifs of the enzyme. However, the results of this study did support some literature speculation that the mechanism of inhibition involved an opening of the hand-shaped structure of the enzyme. Structural changes in the enzyme were also hypothesised to be involved in the enzyme's recognition and response to different nucleic acid substrates. The enzyme was proposed to exist in six basic global conformations, depending on the activation state of the enzyme, the nature of the nucleic acid template and primer bound, and the presence or absence of non-nucleoside inhibitors. Further research in this field will require more crystal structures or computer-generated models, in particular those with the natural RNA primers and templates bound, and structures with both nucleic acids and inhibitors concurrently bound. More detailed study of the non-nucleoside inhibitor binding pocket should advance the characterisation of its flexibility and aid in rational drug design by understanding the intermolecular interactions the pocket is capable of forming to a bound ligand.

5.3 Drug design studies

Investigations into the optimal use of ligand and protein derived structural information for rational computer aided ligand design established the preliminary methodology for the integration of ligand- and structure-based methodologies. Traditional ligand-based pharmacophores were generated using inhibitors that had previously been crystallised with RT, so that their bioactive conformation and unambiguous binding to the active site were known. These pharmacophores were composed of different combinations of a hydrogen bond acceptor, a ring aromatic, a hydrophobic aliphatic and a positive ionisable feature. Selected hypotheses were also merged to form more specific pharmacophores. Individual and merged hypotheses were tested and validated with the inhibitors from crystal structures not used in the pharmacophore generation, however, the results were ambiguous, and this was primarily attributed to the statistical invalidity of the training set used to generate the hypotheses. Database searching with these pharmacophores demonstrated that the merged hypotheses with more features were more effective in returning more manageable hitlists of compounds to be further considered as potential leads. These large hitlists of compounds can now easily be screened with more specific pharmacophores, such as those generated by structure-based design methods, to reduce the current hitlists of several thousand compounds to approximately a couple of hundred. These can then be manually sorted and analysed with the screening data provided by the National Cancer Institute.

A structure-based query/pharmacophore was developed from an activity 'weighted' model of the three-dimensional x-ray structures whose inhibitors were used in the ligand-based pharmacophore. This was used to search the NCI database to identify compounds that fit the structural and conformational requirements of the structure-based pharmacophore. Of the compounds returned from database searching and identified as possible lead compounds to investigate, two had been previously tested by the NCI, and one had been noted as showing anti-viral activity. This served as preliminary validation of the structure-based methodology. The 'superligand' query/phramacophore described all potential binding interactions of residues of the active site to a potential ligand in the pocket. It was converted to a traditional pharmacophore form with the Hypoedit utility in Catalyst, and represented the first step toward the generation of the 'combiphore', an integration of ligand- and structure-based pharmacophores.

The last objective of this research, to fully integrate these pharmacophores and fully exploit the twofold sources of information in one methodology, was not realised at this time, however, this can be achieved readily with only small additional refinement of the individual ligand- and structure-derived pharmacophores necessary. Comparisons of the pharmacophore and the superligand with superimposition and docking studies, as well as databases searching, will be stimulating and potentially rewarding areas of research to investigate. The combiphore, and its constituent pharmacophore and superligand, represent exciting emergent computational tools for the optimal exploitation of ligand and receptor structure information for rational drug design. They will potentially be able to study and consider both existing data about inhibitors known to be active at a binding site, as well as all potential interactions the binding site may be capable of.

Chapter 6

Experimental

6.1 Synthesis

Reagents and solvents were purchased reagent grade and used without further purification, unless otherwise stated. THF, ether and dioxane were dried over KOH and distilled from sodium. DCM was distilled from phosphorus pentoxide. DMF was distilled from $CaSO_4$ under reduced pressure, and glacial acetic acid was distilled from $CuSO_4$. All reactions were performed in standard glassware under an inert atmosphere of nitrogen. Organic solvent extracts were dried with anhydrous $MgSO_4$ and the solvent removed under reduced pressure in a rotary evaporator. Compounds were dried at 10^{-2} Torr. Analytical thin layer chromatography (TLC) was performed on Merck Kieselgel 60 F254 silica on aluminium sheets. Column chromatography was performed using Merck Kieselgel 60 (0.040-0.063 mm, 230-400 mesh, or 0.063-0.200 mm, 70-230 mesh). All chromatographic solvent proportions were volume for volume.

Melting points (mp) were determined on a Reichert hot-stage melting point apparatus. Temperatures were expressed in degrees Celsius (°C) and were uncorrected.

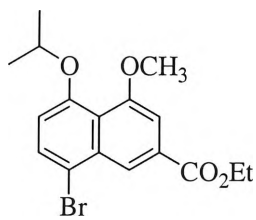
Proton nuclear magnetic resonance (^{1}H NMR) spectra were acquired with Varian Unity-300 or -400 spectrometers at 300.1 and 399.9 MHz respectively. All spectra were recorded using solutions in deuterochloroform, obtained commercially from Aldrich Chemical Company, Inc. or Cambridge Isotope Laboratories, Inc., with chloroform as the internal standard (δ 7.26 ppm). Carbon nuclear magnetic resonance (^{13}C NMR) spectra were acquired with Varian Unity-300 or -400 spectrometers at 75.4 and 100.0 MHz respectively and using deuterochloroform solutions with chloroform as the

internal standard (δ 77.00 ppm). Chemical shifts (δ) and coupling constants (*J*) were expressed in ppm relative to the internal standard, and cycles per second (Hz) respectively. Multiplicities are denoted as s (singlet), d (doublet), bd (broad doublet), dd (doublet of doublets), q (quartet), p (quintet), and m (multiplet). Each resonance was listed according to the following convention: chemical shift, multiplicity, coupling constant, integration, and assignment.

Mass spectra were obtained on a Shimadzu QP-5000 (chemical ionisation, CI), or on a VG Quattro-triple quadropole or MAT-44 quadropole (electrospray, ES) spectrophotometer *via* a direct insertion technique, with an electron beam energy of 70 eV and a source temperature of 200 °C. High resolution mass spectra (HRMS) (CI or ES) were obtained using a Fison/VG Autospec-TOF mass spectrometer.

Microanalyses (elemental analyses) were performed by the Microanalytical Service, Department of Chemistry, The University of Queensland, Brisbane.

Ethyl 8-bromo-5-isopropoxy-4-methoxy-2-naphthalenecarboxylate (65)

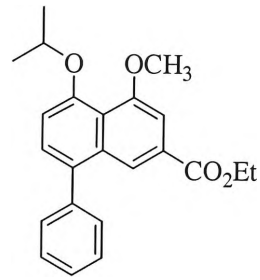


(65)

Compound (**65**) was synthesised in 7 steps according to literature [53,62]. The overall yield for 7 steps of the synthesis to (**65**) was 12.7%. The only modification to the literature was the methylation reaction as detailed following. A mixture of (**64**) (1.00 g, 2.83 mmol), potassium carbonate (1.17 g, 8.47 mmol) and iodomethane (1.76 mL, 28.27 mmol) in dry distilled acetone (200 mL) was heated at reflux under N_2 for 24 hours. The solvent was removed under reduced pressure, and the residue was extracted with chloroform (4 x 50 mL) and water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (10% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**65**) (945 mg, 91%) as an orange solid: mp 66.6-67.4 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.40 (d, *J* = 6.0 Hz, 6H, $CH(CH_3)_2$), 1.45 (t, *J* = 7.0 Hz, 3H, CH_2CH_3), 4.00 (s, 3H, OCH_3), 4.45 (q, *J* = 7.2 Hz, 2H, CH_2CH_3), 4.53 (p, *J* = 6.0 Hz, 1H, $CH(CH_3)_2$), 6.87 (d, *J* = 8.4 Hz, 1H, *H*-6), 7.46 (d, *J* = 1.2 Hz, 1H, *H*-3), 7.70 (d, *J* = 8.4 Hz, 1H, *H*-7), 8.56 (d, *J* = 1.2 Hz, 1H, *H*-1); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.36 (CH_2CH_3), 21.93 ($CH(CH_3)_2$), 56.42 (OCH_3), 61.33 (CH_2CH_3), 73.34 ($CH(CH_3)_2$), 105.84 (*C*-3), 114.82 (*C*-6), 115.52 (*C*-8a), 122.62 (*C*-4a), 122.73 (*C*-1), 129.19 (*C*-2), 131.08 (*C*-7), 134.40 (*C*-8), 154.98 (*C*-5), 157.65 (*C*-4), 166.47 (*C*OO); *m/z* (CI^+) 369 (*M*+1, Br^{81}, 93), 367

(M+1, Br^{79}, 100), 327 (Br^{81}, 12), 325 (Br^{79}, 11), 289 (25), 247 (6); HRMS (CI^+) calcd for $C_{17}H_{20}BrO_4$: 367.0545 found 367.0541.

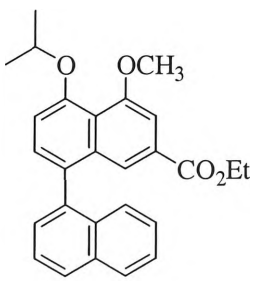
Ethyl 5-isopropoxy-4-methoxy-8-phenyl-2-naphthalenecarboxylate (68)



(68)

A mixture of (**65**) (500 mg, 1.36 mmol), phenylboronic acid (350 mg, 2.87 mmol), barium hydroxide (700 mg, 4.08 mmol) and $PdCl_2$(dppf) (115 mg, 141 μmol) in dry distilled THF (30 mL) was heated at reflux under N_2 for 48 hours. The solvent was removed under reduced pressure, and the residue was partitioned between chloroform and 2M NaOH. The aqueous layer was extracted with chloroform (4 x 50 mL), and the combined organic layers washed with water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (10% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**68**) (480 mg, 97%) as a yellow solid: mp 104.7-106.9 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.34 (t, J = 7.0 Hz, 3H, CH_2CH_3), 1.45 (d, J = 6.4 Hz, 6H, $CH(CH_3)_2$), 4.03 (s, 3H, OCH_3), 4.34 (q, J = 7.2 Hz, 2H, CH_2CH_3), 4.59 (p, J = 6.1 Hz, 1H, $CH(CH_3)_2$), 7.09 (d, J = 8.0 Hz, 1H, *H*-6), 7.36 (d, J = 8.0 Hz, 1H, *H*-7), 7.44 (m, 5H, *aryl*), 7.48 (d, J = 1.6 Hz, 1H, *H*-3), 8.22 (d, J = 1.6 Hz, 1H, *H*-1); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.27 (CH_2CH_3), 22.14 ($CH(CH_3)_2$), 56.46 (OCH_3), 61.00 (CH_2CH_3), 73.33 ($CH(CH_3)_2$), 105.19 (*C*-3), 114.64 (*C*-6), 121.69 (*C*-4a), 122.13 (*C*-1), 127.15 (*C*-7), 127.88 (*C*-2), 128.27 (*C*-2', 6'), 128.36 (*C*-4'), 130.26 (*C*-3', 5'), 134.66 (*C*-1'), 135.17 (*C*-8), 140.81 (*C*-8a), 154.50 (*C*-5), 157.55 (*C*-4), 166.83 (*C*OO); *m/z* (CI^+) 365 (M+1, 100), 323 (16), 263 (16); HRMS (CI^+) calcd for $C_{23}H_{25}O_4$: 365.1753, found 365.1749.

Ethyl 5-isopropoxy-4-methoxy-8-naphthyl-2-naphthalenecarboxylate (72)



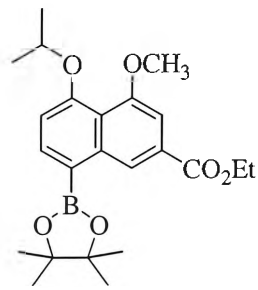
(72)

A mixture of (**65**) (500 mg, 1.36 mmol), naphthylboronic acid (350 mg, 2.73 mmol), potassium acetate (535 mg, 5.45 mmol) and $PdCl_2$(dppf) (110 mg, 135 μmol) in dry distilled THF (20 mL) was heated at reflux in a sealed tube under N_2 for 48 hours. The solvent was removed under reduced pressure, and the residue was partitioned between chloroform and 2M NaOH. The aqueous layer was extracted with chloroform (4 x 50 mL), and

the combined organic layers washed with water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (10% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**72**) (510 mg, 91%) as a yellow solid: mp 166.6-167.3 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.20 (t, *J* = 7.2 Hz, 3H, CH_2CH_3), 1.49 (d, *J* = 6.4 Hz, 6H, $CH(CH_3)_2$), 1.50 (d, *J* = 6.0 Hz, 6H, $CH(CH_3)_2$), 4.05 (s, 3H, OCH_3), 4.21 (q, *J* = 6.8 Hz, 2H, CH_2CH_3), 4.66 (p, *J* = 6.1 Hz, 1H, $CH(CH_3)_2$), 7.14 (d, *J* = 8.0 Hz, 1H, *H*-6), 7.29 (ddd, *J* = 1.3, 7.0, 8.4 Hz, 1H, *naphthyl*), 7.37 (s, 1H, *naphthyl*), 7.39 (d, *J* = 1.6 Hz, 1H, *H*-3), 7.41 (d, *J* = 8.0 Hz, 1H, *H*-7), 7.46 (ddd, *J* = 1.2, 4.0, 4.0 Hz, 1H, *naphthyl*), 7.47 (ddd, *J* = 1.3, 4.1, 4.1 Hz, 1H, *naphthyl*), 7.58 (dd, *J* = 0.8, 7.2 Hz, 1H, *naphthyl*), 7.74 (d, *J* = 1.2 Hz, 1H, *H*-1), 7.94 (dd, *J* = 1.4, 8.2 Hz, 2H, *naphthyl*); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.11 (CH_2CH_3), 22.16 ($CH(CH_3)_2$), 22.22 ($CH(CH_3)_2$), 56.46 (OCH_3), 60.86 (CH_2CH_3), 73.23 ($CH(CH_3)_2$), 105.12 (*C*-3), 114.17 (*C*-6), 121.45, 122.43 (*C*-1), 125.31 (*naphthyl*), 125.71, 125.94, 126.37, 127.93, 128.16, 128.24, 128.57, 129.33, 132.92, 133.14, 133.70, 135.89, 138.27, 154.78 (*C*-5), 157.54 (*C*-4), 166.66 (*C*OO); *m/z* (CI^+) 415 (M^+, 56), 401 (12), 338 (25), 263 (44); HRMS (CI^+) calcd for $C_{27}H_{27}O_4$: 415.1909, found 415.1887.

Ethyl 5-isopropoxy-4-methoxy-8-(4',4',5',5'-tetramethyl-[1,3,2]dioxaborolan-2-yl)-2-naphthalenecarboxylate (73)

A mixture of (**65**) (200 mg, 5.45 mmol), bis(pinacolato)diboron (280 mg, 1.10 mmol), potassium acetate (220 mg, 2.24 mmol) and $PdCl_2$(dppf) (45 mg, 55.10 μmol) in dry distilled THF (25 mL) was heated at reflux under N_2 for 24 hours. The solvent was removed under reduced pressure, and the residue was extracted with chloroform (4 x 50 mL) and water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (10% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**73**) (215 mg, 95%) as a pale yellow solid: mp 98.7-101.3 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.25 (s, 3H, $C(CH_3)_2$), 1.26 (s, 3H, $C(CH_3)_2$), 1.41 (d, *J* = 6.0 Hz, 6H, $CH(CH_3)_2$), 1.42 (s, 3H, $C(CH_3)_2$), 1.42 (s, 3H, $C(CH_3)_2$), 1.45 (t, *J* = 7.2 Hz, 3H, CH_2CH_3), 3.97 (s, 3H, OCH_3), 4.41 (q, *J* = 7.1 Hz, 2H, CH_2CH_3), 4.64 (p, *J* = 6.1 Hz, 1H, $CH(CH_3)_2$), 6.97 (d, *J* = 8.0 Hz, 1H, *H*-6), 7.39 (s, 1H, *H*-3), 7.99 (d, *J* = 8.0 Hz, 1H, *H*-7), 9.19 (d, *J* = 1.2 Hz, 1H, *H*-1); ^{13}C NMR

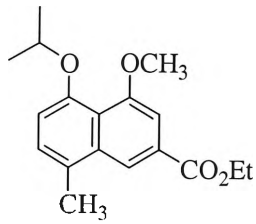


(**73**)

($CDCl_3$, 100 MHz) δ 14.26 (CH_2CH_3), 21.96 ($CH(CH_3)_2$), 24.92 ($C(CH_3)_2$), 24.99 ($C(CH_3)_2$), 56.34 (OCH_3), 60.88 (CH_2CH_3), 72.27 ($CH(CH_3)_2$), 83.47 ($C(CH_3)_2$), 83.65 ($C(CH_3)_2$), 104.97 (*C*-3), 112.36 (*C*-6), 121.07 (*C*-2, 4a), 124.40 (*C*-1), 128.10 (*C*-8a), 137.17 (*C*-7), 140.47 (*C*-8), 157.58 (*C*-4), 157.93 (*C*-5), 167.21 (*C*OO); *m/z* (CI^+) 415 (*M*+1, 100), 373 (13), 289 (5); HRMS (CI^+) calcd for $C_{23}H_{32}BO_6$: 415.2292, found 415.2290.

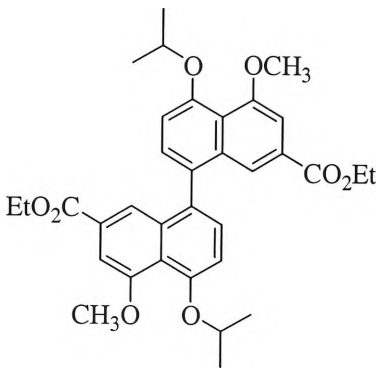
Ethyl 5-isopropoxy-4-methoxy-8-methyl-2-naphthalenecarboxylate (79)

A mixture of (**65**) (400 mg, 1.09 mmol), methylboronic acid (130 mg, 2.17 mmol), potassium carbonate (600 mg, 4.34 mmol) and $PdCl_2$(dppf) (90 mg, 110 μmol) in dry distilled THF (30 mL) was heated at reflux under N_2 for 48 hours. The solvent was removed under reduced pressure, and the residue was partitioned between chloroform and 2M NaOH. The aqueous layer was extracted with chloroform (4 x 50 mL), and the combined organic layers washed with water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (10% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**79**) (311 mg, 94%) as a pale orange solid: mp 207.8-209.1 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.38 (d, *J* = 6.0 Hz, 6H, $CH(CH_3)_2$), 1.44 (t, *J* = 7.2 Hz, 3H, CH_2CH_3), 2.65 (s, 3H, Ar-CH_3), 4.01 (s, 3H, OCH_3), 4.44 (q, *J* = 7.2 Hz, 2H, CH_2CH_3), 4.46 (p, *J* = 6.0 Hz, 1H, $CH(CH_3)_2$), 6.09 (d, *J* = 7.6 Hz, 1H, *H*-6), 7.24 (d, *J* = 8.0 Hz, 1H, *H*-7), 7.43 (d, *J* = 0.8 Hz, 1H, *H*-3), 7.31 (d, *J* = 0.8 Hz, 1H, *H*-1); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 14.32 (CH_2CH_3), 19.75 (Ar-CH_3), 21.94 ($CH(CH_3)_2$), 56.14 (OCH_3), 61.00 (CH_2CH_3), 73.47 ($CH(CH_3)_2$), 104.82 (*C*-3), 115.88 (*C*-6), 119.97 (*C*-1), 121.71 (*C*-4a), 127.45 (*C*-8a), 127.74 (*C*-7), 128.94 (*C*-2), 135.19 (*C*-8), 153.20 (*C*-5), 157.46 (*C*-4), 166.83 (*C*OO); *m/z* (CI^+) 303 (*M*+1, 100), 261 (13); HRMS (CI^+) calcd for $C_{18}H_{23}O_4$: 303.1596, found 303.1518.



(**79**)

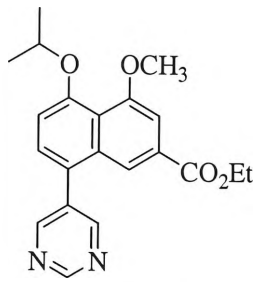
Diethyl 5,5'-diisopropoxy-4,4'-dimethoxy-8,8'-diphenyl-(8,8'-binaphthyl)-2,2'-dicarboxylate (80)



(80)

A mixture of (**65**) (200 mg, 545 μmol), bis(pinacolato)diboron (280 mg, 1.10 mmol), potassium carbonate (300 mg, 2.17 mmol) and $PdCl_2(dppf)$ (50 mg, 61 μmol) in dry distilled THF (30 mL) was heated at reflux in a sealed tube, under N_2. After 24 hours a mixture of methyl iodide (110 μL, 1.77 mmol), barium hydroxide (380 mg, 2.22 mmol), and a second equivalent of $PdCl_2(dppf)$ (50 mg, 61 μmol) was added and heated at reflux under N_2 for a further 24 hours. The solvent was removed under reduced pressure, and the residue was partitioned between chloroform and 2M NaOH. The aqueous layer was extracted with chloroform (4 x 50 mL), and the combined organic layers washed with water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (20% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**80**) (144 mg, 92%) as an orange solid: mp 173.1-174.6 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.21 (t, *J* = 7.2 Hz, 6H, CH_2CH_3, $CH'_2CH'_3$), 1.47 (d, *J* = 6.0 Hz, 6H, $CH(CH_3)_2$, $CH'(CH'_3)_2$), 1.49 (d, *J* = 6.4 Hz, 6H, $CH(CH_3)_2$, $CH'(CH'_3)_2$), 4.04 (s, 6H, OCH_3, OCH'_3), 4.21 (q, *J* = 7.2 Hz, 2H, CH_2CH_3, $CH'_2CH'_3$), 4.22 (q, *J* = 7.2 Hz, 2H, CH_2CH_3, $CH'_2CH'_3$), 4.65 (p, *J* = 6.0 Hz, 2H, $CH(CH_3)_2$, $CH'(CH'_3)_2$), 7.13 (d, *J* = 8.0 Hz, 2H, *H*-6, *H*-6'), 7.37 (d, *J* = 7.6 Hz, 2H, *H*-7, *H*-7'), 7.38 (d, *J* = 1.6 Hz, 2H, *H*-3, *H*-3'), 7.68 (d, *J* = 1.6 Hz, 2H, *H*-1, *H*-1'); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.12 (CH_2CH_3, $CH'_2CH'_3$), 22.09 ($CH(CH_3)_2$, $CH'(CH'_3)_2$), 22.17 ($CH(CH_3)_2$, $CH'(CH'_3)_2$), 56.38 (OCH_3, OCH'_3), 60.85 (CH_2CH_3, $CH'_2CH'_3$), 73.27 ($CH(CH_3)_2$, $CH'(CH'_3)_2$), 105.01 (*C*-3, *C*-3'), 114.43 (*C*-6, *C*-6'), 121.62 (*C*-4a, *C*-4a'), 122.23 (*C*-1, *C*-1'), 127.85 (*C*-2, *C*-2'), 129.68 (*C*-7, *C*-7'), 132.80 (*C*-8, *C*-8'), 136.15 (*C*-8a, *C*-8a'), 154.68 (*C*-5, *C*-5'), 157.46 (*C*-4, *C*-4'), 166.66 (*C*OO, *C*'OO); *m/z* (ES^+) 575 (*M*+1, 38), 529 (43), 487 (8); HRMS (ES^+) calcd for $C_{34}H_{39}O_8$: 575.2645, found 575.2647.

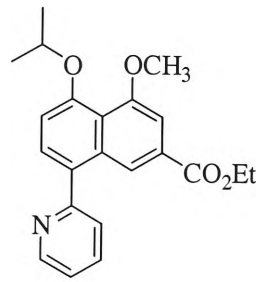
Ethyl 5-isopropoxy-4-methoxy-8-(5'-pyrimidyl)-2-naphthalenecarboxylate (83)



(83)

A mixture of (**65**) (200 mg, 0.54 mmol), bis(pinacolato)diboron (280 mg, 1.10 mmol), potassium acetate (220 mg, 2.24 mmol) and $PdCl_2$(dppf) (45 mg, 0.06 mmol) in dry distilled THF (30 mL) was heated at reflux in a sealed tube, under N_2. After 24 hours a mixture of 5-bromopyrimidine (260 mg, 1.64 mmol), barium hydroxide (375 mg, 2.19 mmol), and a second equivalent of $PdCl_2$(dppf) (45 mg, 0.06 mmol) was added and heated at reflux under N_2 for a further 24 hours. The solvent was removed under reduced pressure, and the residue was partitioned between chloroform and 2M NaOH. The aqueous layer was extracted with chloroform (4 x 50 mL), and the combined organic layers washed with water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (40% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**83**) (154 mg, 77%) as a yellow solid: mp 175.7-177.9 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.34 (t, J = 7.2 Hz, 3H, CH_2CH_3), 1.45 (d, J = 6.0 Hz, 6H, $CH(CH_3)_2$), 4.02 (s, 3H, OCH_3), 4.35 (q, J = 7.2 Hz, 2H, CH_2CH_3), 4.64 (p, J = 6.0 Hz, 1H, $CH(CH_3)_2$), 7.07 (d, J = 8.4 Hz, 1H, *H*-6), 7.32 (d, J = 8.0 Hz, 1H, *H*-7), 7.44 (d, J = 0.8 Hz, 1H, *H*-3), 7.96 (d, J = 1.2 Hz, 1H, *H*-1), 8.83 (s, 2H, *H*-4', 6'), 9.27 (s, 1H, *H*-2'); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 14.24 (CH_2CH_3), 21.98 ($CH(CH_3)_2$), 56.37 (OCH_3), 61.25 (CH_2CH_3), 72.89 ($CH(CH_3)_2$), 105.62 (*C*-3), 112.98 (*C*-6), 119.88 (*C*-1), 121.38 (*C*-8a), 126.43 (*C*-4a), 128.99 (*C*-2), 129.32 (*C*-7), 134.34 (*C*-1'), 134.56 (*C*-8), 156.13 (*C*-5), 157.41 (*C*-2', *C*-4'), 157.98 (*C*-4), 166.28 (*C*OO); *m/z* (CI^+) 367 (*M*+1, 100), 338 (4), 325 (6); HRMS (CI^+) calcd for $C_{21}H_{23}N_2O_4$: 367.1658, found 367.1648. (MA found: C, 69.06; H, 6.00; N, 7.72; O, 17.22; $C_{21}H_{22}N_2O_4$ requires C, 68.84; H, 6.05; N, 7.65; O, 17.47).

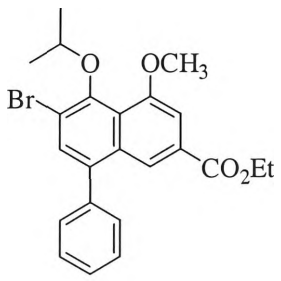
Ethyl 5-isopropoxy-4-methoxy-8-(2'-pyridyl)-2-naphthalenecarboxylate (84)



(**84**)

A mixture of (**73**) (225 mg, 543 μmol), 2-bromopyrimidine (105 μL, 1.10 mmol), potassium carbonate (225 mg, 1.63 mmol), triphenylphosphine (145 mg, 553 mmol) and palladium acetate (15 mg, 67 μmol) in DME (15 mL) was heated at reflux in a sealed tube, under N_2, for 24 hours. The solvent was removed under reduced pressure, and the residue was partitioned between chloroform and 2M NaOH. The aqueous layer was extracted with chloroform (4 x 25 mL), and the combined organic layers washed with water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (35% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**84**) (90 mg, 45%) as a pale yellow solid: mp 71.4-72.7 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.36 (t, *J* = 7.2 Hz, 3H, CH_2CH_3), 1.44 (d, *J* = 6.0 Hz, 6H, $CH(CH_3)_2$), 4.02 (s, 3H, OCH_3), 4.36 (q, *J* = 7.1 Hz, 2H, CH_2CH_3), 4.62 (p, *J* = 6.1 Hz, 1H, $CH(CH_3)_2$), 7.11 (d, *J* = 8.0 Hz, 1H, *H*-6), 7.31 (ddd, *J* = 0.8, 7.4, 7.4 Hz, 1H, *H*-5'), 7.42 (d, *J* = 1.2 Hz, 1H, *H*-3), 7.54 (d, *J* = 7.6 Hz, 1H, *H*-3'), 7.54 (d, *J* = 8.4 Hz, 1H, *H*-7), 7.81 (ddd, *J* = 1.6, 7.6, 7.6 Hz, 1H, *H*-4'), 8.39 (d, *J* = 1.2 Hz, 1H, *H*-1), 8.78 (bd, *J* = 4.0 Hz, 1H, *H*-6'); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 14.24 (CH_2CH_3), 21.99 ($CH(CH_3)_2$), 56.34 (OCH_3), 60.99 (CH_2CH_3), 73.13 ($CH(CH_3)_2$), 105.17 (*C*-3), 114.13 (*C*-6), 121.53 (*C*-1), 121.61 (*C*-4a), 121.86 (*C*-5'), 125.17 (*C*-7), 128.24 (*C*-2), 129.06 (*C*-3'), 133.25 (*C*-8), 134.23 (*C*-8a), 136.35 (*C*-4'), 149.50 (*C*-6'), 155.43 (*C*-5), 157.54 (*C*-4), 159.11 (*C*-2'), 166.76 (*C*OO); *m/z* (CI^+) 366 (*M*+1, 100), 324 (10); HRMS (CI^+) calcd for $C_{22}H_{24}NO_4$: 366.1705, found 366.1697.

Ethyl 6-bromo-5-isopropoxy-4-methoxy-8-phenyl-2-naphthalenecarboxylate (85)

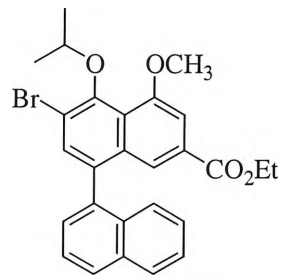


(**85**)

To a solution of (**68**) (365 mg, 1.00 mmol) in dry chloroform (60 mL) was added dropwise pyridinium tribromide (355 mg, 1.11 mmol) in distilled glacial acetic acid (15 mL) under N_2, at 0 °C. The resulting solution was allowed to stir at room temperature for 3 hours followed by the addition of saturated $NaHCO_3$ (250 mL). The aqueous layer was extracted with chloroform (3 x 50 mL), and the combined organic layers were

washed with water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before chromatography with a short silica plug (10% ethyl acetate:hexane). Recrystallisation from isopropanol afforded (**85**) (340 mg, 76%) as an orange solid: mp 132.4-133.5 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.34 (t, *J* = 7.0 Hz, 3H, CH_2CH_3), 1.38 (d, *J* = 6.0 Hz, 6H, $CH(CH_3)_2$), 4.06 (s, 3H, OCH_3), 4.35 (q, *J* = 7.0 Hz, 2H, CH_2CH_3), 4.56 (p, *J* = 6.2 Hz, 1H, $CH(CH_3)_2$), 7.43 (d, *J* = 1.6 Hz, 1H, *H*-3), 7.47 (m, 5H, *aryl*), 7.65 (s, 1H, *H*-7), 8.17 (d, *J* = 1.2 Hz, 1H, *H*-1); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.23 (CH_2CH_3), 22.04 ($CH(CH_3)_2$), 56.04 (OCH_3), 61.13 (CH_2CH_3), 78.12 ($CH(CH_3)_2$), 105.48 (*C*-3), 118.19 (*C*-6), 122.26 (*C*-1), 124.08 (*C*-4a), 127.75 (*C*-4'), 128.09 (*C*-2), 128.38 (*C*-3'), 130.05 (*C*-2'), 132.51 (*C*-7), 138.82 (*C*-8a), 138.22 (*C*-1'), 139.15 (*C*-8), 150.17 (*C*-5), 155.98 (*C*-4), 166.43 (*C*OO); *m/z* (CI^+) 445 (*M*+1, Br^{81}, 67), 443 (*M*+1, Br^{79}, 70), 429 (74), 401 (43) 387 (47); HRMS (CI^+) calcd for $C_{23}H_{24}BrO_4$: 443.0858, found 443.0856.

Ethyl 6-bromo-5-isopropoxy-4-methoxy-8-naphthyl-2-naphthalenecarboxylate (86)

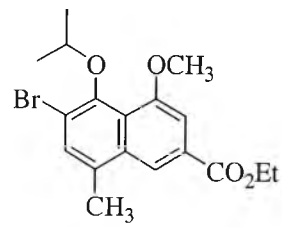


(**86**)

To a solution of (**72**) (230 mg, 555 μmol) in dry chloroform (60 mL) was added dropwise pyridinium tribromide (195 mg, 610 μmol) in distilled glacial acetic acid (15 mL) under N_2, at 0 °C. The resulting solution was allowed to stir at room temperature for 3 hours followed by the addition of saturated $NaHCO_3$ (250 mL). The aqueous layer was extracted with chloroform (3 x 50 mL), and the combined organic layers were washed with water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before chromatography with a short silica plug (10% ethyl acetate:hexane). Recrystallisation from isopropanol afforded (**86**) (266 mg, 97%) as an orange solid: mp 142.8-145.0 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.20 (t, *J* = 7.2 Hz, 3H, CH_2CH_3), 1.43 (d, *J* = 6.0 Hz, 6H, $CH(CH_3)_2$), 4.08 (s, 3H, OCH_3), 4.22 (q, *J* = 7.2 Hz, 2H, CH_2CH_3), 4.64 (p, *J* = 6.1 Hz, 1H, $CH(CH_3)_2$), 7.33 (m, 2H, *naphthyl*), 7.43 (d, *J* = 1.6 Hz, 1H, *H*-3), 7.47 (s, 1H, *H*-7), 7.48 (m, 3H, *naphthyl*), 7.70 (d, *J* = 1.2 Hz, 1H, *H*-1), 7.95 (m, *J* = 8.4 Hz, 2H, *naphthyl*); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 14.11 (CH_2CH_3), 22.10 ($CH(CH_3)_2$), 22.13 ($CH(CH_3)_2$), 56.06 (OCH_3), 61.03 (CH_2CH_3), 78.21 ($CH(CH_3)_2$), 105.40 (*C*-3), 118.13 (*C*-6), 122.53 (*C*-1), 123.91 (*naphthyl*), 125.24, 125.92, 126.00, 126.27, 128.16, 128.26, 128.49, 132.69, 133.44, 133.62, 135.05, 136.33, 136.53 (*C*-8), 150.46 (*naphthyl*), 154.33 (*C*-5), 155.93 (*C*-4), 166.28 (*C*OO); *m/z* (CI^+) 495 (*M*+1, Br^{81}, 95), 493 (*M*+1,

Br^{79}, 100), 453 (Br^{81}, 64), 451 (Br^{79}, 62); HRMS (CI^+) calcd for $C_{27}H_{26}BrO_4$: 493.1014, found 493.1020.

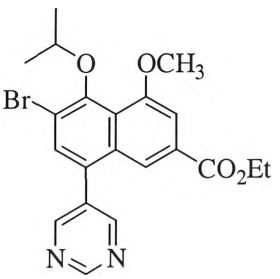
Ethyl 6-bromo-5-isopropoxy-4-methoxy-8-methyl-2-naphthalenecarboxylate (87)



(87)

To a solution of (**79**) (150 mg, 0.50 mmol) in dry chloroform (60 mL) was added dropwise pyridinium tribromide ((175 mg, 0.55 mmol) in distilled glacial acetic acid (15 mL) under N_2, at 0 °C. The resulting solution was allowed to stir at room temperature for 3 hours followed by the addition of saturated $NaHCO_3$ (250 mL). The aqueous layer was extracted with chloroform (3 x 50 mL), and the combined organic layers were washed with water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before chromatography with a short silica plug (20% ethyl acetate:hexane). Recrystallisation from isopropanol afforded (**87**) (172 mg, 91%) as a white solid: mp 88.1-89.2 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.32 (d, *J* = 6.0 Hz, 6H, $CH(CH_3)_2$), 1.44 (t, *J* = 7.2 Hz, 3H, CH_2CH_3), 2.64 (s, 3H, Ar-CH_3), 4.03 (s, 3H, OCH_3), 4.45 (q, *J* = 7.2 Hz, 2H, CH_2CH_3), 4.47 (p, *J* = 6.0 Hz, 1H, $CH(CH_3)_2$), 7.46 (d, *J* = 1.6 Hz, 1H, *H*-3), 7.54 (s, 1H, *H*-7), 8.26 (d, *J* = 1.6 Hz, 1H, *H*-1); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 14.39 (CH_2CH_3), 19.59 (Ar-CH_3), 21.92 ($CH(CH_3)_2$), 55.93 (OCH_3), 61.25 (CH_2CH_3), 77.78 ($CH(CH_3)_2$), 105.39 (*C*-3), 118.19 (*C*-6), 120.07 (*C*-1), 123.85 (*C*-4a), 127.83 (*C*-8a), 132.16 (*C*-7), 132.27 (*C*-2), 134.46 (*C*-8), 149.07 (*C*-5), 156.12 (*C*-4), 166.58 (*C*OO); *m/z* (CI^+) 383 (*M*+1, Br^{81}, 91), 381 (*M*+1, Br^{79}, 96), 339 (Br^{81}, 100), 341 (*M*+1, Br^{79}, 92), 302 (61), 261 (56); HRMS (CI^+) calcd for $C_{18}H_{22}BrO_4$: 381.0701, found 381.0693.

Ethyl 6-bromo-5-hydroxy-4-methoxy-8-(5'-pyrimidyl)-2-naphthalenecarboxylate (88)



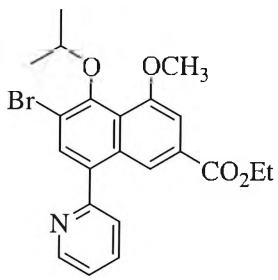
(88)

To a solution of (**83**) (130 mg, 355 μmol) in dry chloroform (40 mL) and distilled glacial acetic acid (10 mL), was added dropwise pyridinium tribromide (125 mg, 391 μmol) in distilled glacial acetic acid (15 mL) under N_2, at 0 °C. The resulting solution was allowed to stir at room temperature for 3 hours followed by the addition of saturated $NaHCO_3$ (250 mL). The aqueous layer was extracted with chloroform (3 x 50 mL), and

the combined organic layers were washed with water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before chromatography with a short silica plug (40% ethyl acetate:hexane). Recrystallisation from isopropanol afforded (**88**) (155 mg, 98%) as a light orange solid: mp 201.8-203.4 °C; ^{1}H NMR ($CDCl_3$, 300 MHz) δ 1.35 (t, J = 7.2 Hz, 3H, CH_2CH_3), 1.39 (d, J = 6.0 Hz, 6H, $CH(CH_3)_2$), 4.08 (s, 3H, OCH_3), 4.37 (q, J = 7.1 Hz, 2H, CH_2CH_3), 4.58 (p, J = 6.1 Hz, 1H, $CH(CH_3)_2$), 7.52 (d, J = 1.2 Hz, 1H, *H*-3), 7.64 (s, 1H, *H*-7), 7.93 (d, J = 1.5 Hz, 1H, *H*-1), 8.85 (s, 2H, *H*-4', 6'), 9.33 (s, 1H, *H*-2'); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.29 (CH_2CH_3), 22.08 ($CH(CH_3)_2$), 56.14 (OCH_3), 61.50 (CH_2CH_3), 78.69 ($CH(CH_3)_2$), 106.11 (*C*-3), 118.07 (*C*-6), 120.31 (*C*-1), 124.18 (*C*-4a), 129.27 (*C*-2), 130.23 (*C*-8a), 133.16 (*C*-1'), 133.32 (*C*-7), 133.57 (*C*-8), 151.91 (*C*-5), 156.30 (*C*-4), 157.32 (*C*-2', 4'), 158.05 (*C*-2'), 165.99 (*C*OO); *m/z* (CI^+) 447 (*M*+1, Br^{81}, 59), 445 (*M*+1, Br^{79}, 54), 405 (Br^{81}, 24), 403 (Br^{79}, 25), 365 (33); HRMS (CI^+) calcd for $C_{21}H_{22}BrN_2O_4$: 445.0727, found 445.0749.

Ethyl 6-bromo-5-hydroxy-4-methoxy-8-(5'-pyridyl)-2-naphthalenecarboxylate (89)

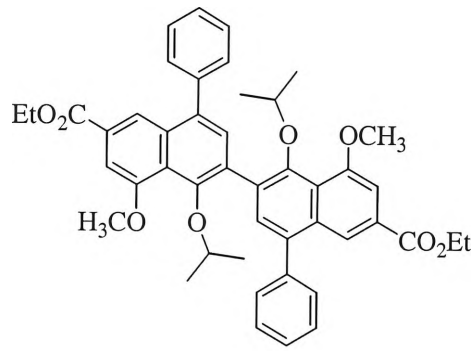
To a solution of (**84**) (85 mg, 233 μmol) in dry chloroform (30 mL) and distilled glacial acetic acid (5 mL), was added dropwise pyridinium tribromide (82 mg, 256 μmol) in distilled glacial acetic acid (10 mL) under N_2, at 0 °C. The resulting solution was allowed to stir at room temperature for 3 hours followed by the addition of saturated $NaHCO_3$ (100 mL). The aqueous layer was extracted with chloroform (3 x 50 mL), and the combined organic layers were washed with water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before chromatography with a short silica plug (35% ethyl acetate:hexane). Recrystallisation from isopropanol afforded (**89**) (83 mg, 81%) as an orange solid: mp 94.2-97.8 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.35 (t, J = 7.2 Hz, 3H, CH_2CH_3), 1.37 (d, J = 6.0 Hz, 6H, $CH(CH_3)_2$), 4.05 (s, 3H, OCH_3), 4.35 (q, J = 7.1 Hz, 2H, CH_2CH_3), 4.55 (p, J = 6.1 Hz, 1H, $CH(CH_3)_2$), 7.37 (ddd, J = 0.8, 5.0, 7.4 Hz, 1H, *H*-5'), 7.46 (d, J = 1.2 Hz, 1H, *H*-3), 7.55 (d, J = 8.0 Hz, 1H, *H*-3'), 7.80 (s, 1H, *H*-7), 7.84 (ddd, J = 1.6, 7.6, 8.4 Hz, 1H, *H*-4'), 8.31 (d, J = 1.2 Hz, 1H, *H*-1), 8.80 (d, J = 4.4 Hz, 1H, *H*-6'); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 14.26 (CH_2CH_3), 22.00 ($CH(CH_3)_2$), 56.01 (OCH_3), 61.18 (CH_2CH_3), 78.39 ($CH(CH_3)_2$), 105.49 (*C*-3), 118.17 (*C*-6), 121.79 (*C*-1), 122.52 (*C*-5'), 124.15 (*C*-4a), 125.18 (*C*-7), 128.48 (*C*-2), 133.12 (*C*-3'), 133.40 (*C*-8), 134.36 (*C*-8a), 136.66 (*C*-4'), 149.65 (*C*-6'), 151.21 (*C*-5), 155.94



(**89**)

(*C*-4), 157.57 (*C*-2'), 166.42 (*C*OO); *m/z* (CI^+) 446 (*M*+1, Br^{81}, 100), 444 (*M*+1, Br^{79}, 84), 404 (Br^{81}, 13), 402 (*M*+1, Br^{79}, 15), 366 (Br^{81}, 21), 364 (Br^{79}, 27), 324 (33); HRMS (CI^+) calcd for $C_{22}H_{23}BrNO_4$: 444.0810, found 444.0805.

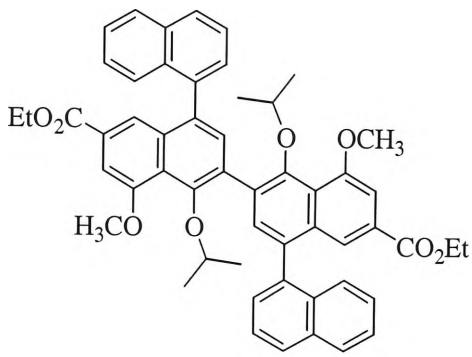
Diethyl 5,5'-diisopropoxy-4,4'-dimethoxy-8,8'-diphenyl-(6,6'-binaphthyl)-2,2'-dicarboxylate (91)



(**91**)

A mixture of (**85**) (100 mg, 226 μmol), bis(pinacolato)diboron (115 mg, 453 μmol), potassium carbonate (125 mg, 904 μmol) and $PdCl_2$(dppf) (20 mg, 25 μmol) in dry distilled THF (25 mL) was heated at reflux in a sealed tube under N_2. After 24 hours a second equivalent of (**85**) (100 mg, 226 μmol), potassium carbonate (125 mg, 904 μmol) and $PdCl_2$(dppf) (20 mg, 25 μmol) was added and heated at reflux under N_2 for a further 24 hours. The solvent was removed under reduced pressure, and the residue extracted with chloroform (4 x 30 mL) and water. The organic phase was dried ($MgSO_4$) and concentrated *in vacuo*, before being subjected to flash column chromatography (20% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**91**) (82 mg, 50%) as a yellow solid: mp 204.8-206.7 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.11 (d, *J* = 5.2 Hz, 12H, $CH(CH_3)_2$, $CH'(CH'_3)_2$), 1.35 (t, *J* = 7.0 Hz, 6H, CH_2CH_3, $CH'_2CH'_3$), 4.06 (s, 6H, OCH_3, OCH'_3), 4.11 (p, *J* = 6.1 Hz, 2H, $CH(CH_3)_2$, $CH'(CH'_3)_2$), 4.36 (q, *J* = 7.0 Hz, 4H, CH_2CH_3, $CH'_2CH'_3$), 7.45 (s, 10H, *aryl''*, *aryl'''*), 7.47 (s, 2H, *H*-3, *H*-3'), 7.96 (s, 2H, *H*-7, *H*-7'), 8.22 (d, *J* = 1.6 Hz, 2H, *H*-1, *H*-1'); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.29 (CH_2CH_3, $C'H_2C'H_3$), 21.88 ($CH(CH_3)_2$, $C'H(C'H_3)_2$), 55.71 (OCH_3, $OC'H_3$), 61.05 (CH_2CH_3, $C'H_2C'H_3$), 74.80 ($CH(CH_3)_2$, $C'H(C'H_3)_2$), 104.82 (*C*-3, *C*-3'), 122.02 (*C*-1, *C*-1'), 124.58 (*C*-4a, *C*-4a'), 127.29 (*C*-4'', *C*-4'''), 127.87 (*C*-2, *C*-2'), 128.21 (*C*-2'', *C*-6'', *C*-2''', *C*-6'''), 130.18 (*C*-3'', *C*-5'', *C*-3''', *C*-5'''), 130.83 (*C*-6, *C*-6'), 131.96 (*C*-7, *C*-7'), 134.06 (*C*-1'', *C*-1'''), 137.08 (*C*-8, *C*-8'), 140.28 (*C*-8a, *C*-8a'), 150.66 (*C*-5, *C*-5'), 157.42 (*C*-4, *C*-4'), 166.76 (*C*OO, *C*'OO); *m/z* (ES^+) 727 (*M*+1, 100), 685 (26), 643 (32); HRMS (ES^+) calcd for $C_{46}H_{47}O_8$: 727.3271, found 727.3296.

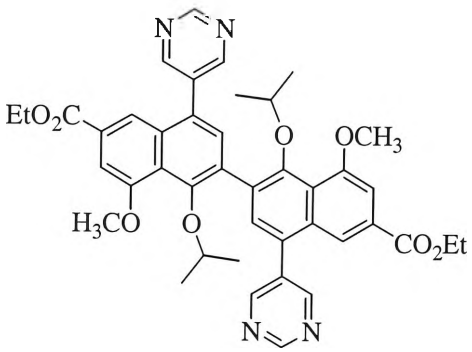
Diethyl 5,5'-diisopropoxy-4,4'-dimethoxy-8,8'-(1'',1'''-dinaphthyl)-(6,6'-binaphthyl)-2,2'-dicarboxylate (96)



(**96**)

A mixture of (**86**) (100 mg, 203 μmol), bis(pinacolato)diboron (105 mg, 413 μmol), potassium carbonate (115 mg, 832 μmol) and $PdCl_2$(dppf) (18 mg, 22 μmol) in dry distilled THF (25 mL) was heated at reflux in a sealed tube under N_2. After 24 hours a second equivalent of (**86**) (100 mg, 203 μmol), potassium carbonate (115 mg, 832 μmol) and $PdCl_2$(dppf) (18 mg, 22 μmol) was added and heated at reflux under N_2 for a further 24 hours. The solvent was removed under reduced pressure, and the residue extracted with chloroform (4 x 30 mL) and water. The organic phase was dried ($MgSO_4$) and concentrated *in vacuo*, before being subjected to flash column chromatography (20% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**96**) (85 mg, 51%) as a yellow solid: mp 286.7-289.3 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.13 (d, *J* = 6.4 Hz, 12H, CH(C*H*$_3$)$_2$, CH'(C*H'*$_3$)$_2$), 1.20 (t, *J* = 7.2 Hz, 6H, CH_2CH_3, $CH'_2CH'_3$), 4.03 (s, 6H, OC*H*$_3$, OC*H'*$_3$), 4.21 (q, *J* = 7.2 Hz, 4H, CH_2CH_3, $CH'_2CH'_3$), 4.24 (p, *J* = 6.8 Hz, 2H, C*H*(CH$_3$)$_2$, C*H'*(CH'$_3$)$_2$), 7.30 (t, *J* = 7.4 Hz, 2H, *naphthyl''*, *naphthyl'''*), 7.40 (d, *J* = 1.2 Hz, 2H, *H*-3, *H*-3'), 7.46 (m, 6H, *naphthyl''*, *naphthyl'''*), 7.57 (t, *J* = 7.6 Hz, 2H, *naphthyl''*, *naphthyl'''*), 7.74 (s, 2H, *H*-1, *H*-1'), 7.93 (dd, *J* = 2.0, 8.0 Hz, 4H, *naphthyl''*, *naphthyl'''*), 8.04 (s, 2H, *H*-7, *H*-7'); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.14 (CH_2CH_3, $CH'_2CH'_3$), 22.11 (CH(*C*H$_3$)$_2$, CH'(*C'*H$_3$)$_2$), 55.73 (OCH$_3$, OC'H$_3$), 60.91 (*C*H$_2$CH$_3$, *C'*H$_2$C'H$_3$), 75.09 (*C*H(CH$_3$)$_2$, *C'*H(C'H$_3$)$_2$), 104.83 (*C*-3, *C*-3'), 122.30 (*C*-1, *C*-1'), 124.40 (*naphthyl''*, *naphthyl'''*), 125.27, 125.69, 125.95, 126.41, 127.99, 128.03, 128.09, 128.15, 130.90, 132.85 (*C*-7, *C*-7'), 132.99 (*naphthyl''*, *naphthyl'''*), 133.59, 135.13, 135.24, 137.84, 151.00 (*C*-5, *C*-5'), 157.36 (*C*-4, *C*-4'), 166.58 (*C*OO, *C'*OO); *m/z* (ES$^+$) 827 (*M*+1, 37), 391 (100); HRMS (ES$^+$) calcd for $C_{54}H_{51}O_8$: 827.3583, found 827.3953.

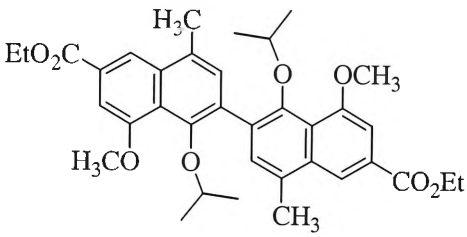
Diethyl 5,5'-diisopropoxy-4,4'-dimethoxy-8,8'-(5'',5'''-dipyrimidyl)-(6,6'-binaphthyl)-2,2'-dicarboxylate (97)



(**97**)

A mixture of (**88**) (20 mg, 45 μmol), bis(pinacolato)diboron (25 mg, 98 μmol), potassium carbonate (25 mg, 181 μmol) and $PdCl_2$(dppf) (4 mg, 5 μmol) in dry distilled THF (25 mL) was heated at reflux in a sealed tube under N_2. After 24 hours a second equivalent of (**88**) (20 mg, 45 μmol), potassium carbonate (25 mg, 181 μmol) and $PdCl_2$(dppf) (4 mg, 5 μmol) were added and heated at reflux under N_2 for a further 24 hours. The solvent was removed under reduced pressure, and the residue extracted with chloroform (4 x 20 mL) and water. The organic phase was dried ($MgSO_4$) and concentrated *in vacuo*, before being subjected to flash column chromatography (30% ethyl acetate:hexane to remove trace starting material; 80% ethyl acetate:hexane to elute product). Recrystallisation from isopropanol yielded (**97**) (23 mg, 71%) as a yellow solid: mp 227.5-229.8 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.12 (d, *J* = 6.0 Hz, 12H, CH(C*H*$_3$)$_2$, CH'(C*H'*$_3$)$_2$), 1.36 (t, *J* = 7.2 Hz, 6H, CH_2CH_3, $CH'_2CH'_3$), 1.36 (t, *J* = 7.2 Hz, 6H, CH_2CH_3, $CH'_2CH'_3$), 4.08 (s, 6H, OC*H*$_3$, OC*H'*$_3$), 4.15 (p, *J* = 6.0 Hz, 2H, C*H*(CH$_3$)$_2$, C*H'*(CH'$_3$)$_2$), 4.380 (q, *J* = 7.1 Hz, 4H, CH_2CH_3, $CH'_2CH'_3$), 4.382 (q, *J* = 7.2 Hz, 4H, CH_2CH_3, $CH'_2CH'_3$), 7.53 (s, 2H, *H*-3, *H*-3'), 7.97 (s, 4H, *H*-1, *H*-1', *H*-7, *H*-7'), 8.88 (s, 4H, *H*-4'', *H*-4''', *H*-5'', *H*-5'''), 9.31 (s, 2H, *H*-2'', *H*-2'''); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.31 (CH_2CH_3, $CH'_2CH'_3$), 21.95 (CH(*C*H$_3$)$_2$, CH'(*C'*H$_3$)$_2$), 55.84 (OCH$_3$, OC'H$_3$), 61.42 (CH_2CH_3, $CH'_2CH'_3$), 75.43 (*C*H(CH$_3$)$_2$, *C'*H(C'H$_3$)$_2$), 105.72 (*C*-3, *C*-3'), 120.03 (*C*-1, *C*-1'), 124.72 (*C*-8a, *C*-8a'), 129.17 (*C*-8, *C*-8'), 129.31 (*C*-4a, *C*-4a'), 130.18 (*C*-6, *C*-6'), 132.65 (*C*-7, *C*-7'), 133.88 (*C*-2, *C*-2'), 134.11 (*C*-5'', *C*-5'''), 152.26 (*C*-5, *C*-5'), 157.36 (*C*-4'', *C*-4''', *C*-6'', *C*-6'''), 157.71 (*C*-4, *C*-4'), 157.76 (*C*-2'', *C*-2'''), 166.16 (*C*OO, *C'*OO); *m/z* (ES$^+$) 731 (*M*+1, 92), 387 (100); HRMS (ES$^+$) calcd for $C_{42}H_{43}N_4O_8$: 731.3081, found 731.3078.

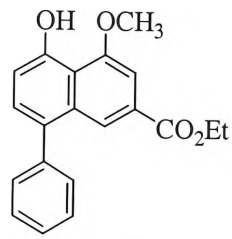
Diethyl 5,5'-diisopropoxy-4,4'-dimethoxy-8,8'-dimethyl-(6,6'-binaphthyl)-2,2'-dicarboxylate (98)



(98)

A mixture of (**87**) (160 mg, 420 μmol), bis(pinacolato)diboron (215 mg, 850 μmol), potassium carbonate (235 mg, 1.70 mmol) and $PdCl_2$(dppf) (35 mg, 43 μmol) in dry distilled THF (25 mL) was heated at reflux in a sealed tube under N_2. After 24 hours, a second equivalent of (**87**) (160 mg, 420 μmol), potassium carbonate (235 mg, 1.70 mmol) and $PdCl_2$(dppf) (35 mg, 43 μmol) was added and heated at reflux under N_2 for a further 24 hours. The solvent was removed under reduced pressure, and the residue extracted with chloroform (4 x 30 mL) and water. The organic phase was dried ($MgSO_4$) and concentrated *in vacuo*, before being subjected to flash column chromatography (10% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**98**) (128 mg, 51%) as a yellow solid: mp 164.8-166.9 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.03 (d, J = 5.2 Hz, 12H, CH(C*H*$_3$)$_2$, CH'(C*H'*$_3$)$_2$), 1.46 (t, J = 7.2 Hz, 6H, CH_2CH_3, $CH'_2CH'_3$), 2.70 (s, 6H, Ar-C*H*$_3$, Ar'-C*H'*$_3$), 3.93 (p, J = 6.0 Hz, 2H, C*H*(CH$_3$)$_2$, C*H'*(CH'$_3$)$_2$), 4.06 (s, 6H, OC*H*$_3$, OC*H'*$_3$), 4.47 (q, J = 7.1 Hz, 4H, CH_2CH_3, $CH'_2CH'_3$), 7.48 (s, 2H, *H*-3, *H*-3'), 7.84 (s, 2H, *H*-7, *H*-7'), 8.34 (d, J = 1.2 Hz, 2H, *H*-1, *H*-1'); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 14.47 (CH_2CH_3, $CH'_2CH'_3$), 19.65 (Ar-CH_3, Ar'-CH'_3), 29.69 (CH(CH_3)$_2$, CH(CH'_3)$_2$), 55.62 (OCH_3, OCH'_3), 61.18 (CH_2CH_3, $CH'_2CH'_3$), 74.41 (CH(CH_3)$_2$, CH(CH'_3)$_2$), 104.76 (*C*-3, *C*-3'), 119.89 (*C*-1, *C*-1'), 124.53 (*C*-6, *C*-6'), 127.59 (*C*-4a, *C*-4a'), 130.55 (*C*-8a, *C*-8a'), 131.00 (*C*-2, *C*-2'), 131.50 (*C*-7, *C*-7'), 134.67 (*C*-8, *C*-8'), 149.51 (*C*-5, *C*-5'), 156.65 (*C*-4, *C*-4'), 166.95 (*C*OO, *C'*OO); *m/z* (CI$^+$) 603 (*M*+1, 42), 589 (16) 561 (21), 519 (30); HRMS (CI$^+$) calcd for $C_{36}H_{43}O_8$: 603.2958, found 603.2953.

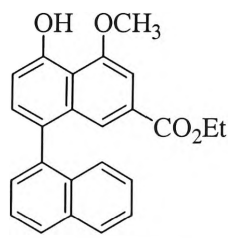
Ethyl 5-hydroxy-4-methoxy-8-phenyl-2-naphthalenecarboxylate (99)



(**99**)

To a solution of (**68**) (95 mg, 261 μmol) in DCM (20 mL) was added dropwise boron trifluoride diethyl etherate (125μL, 1.06 mmol) under N_2, at 0°C. The resulting solution was allowed to stir at room temperature for 24 hours, followed by the addition of methanol (5 mL). The solvent was removed under reduced pressure, and the residue extracted with chloroform (4 x 20 mL) and water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (20% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**99**) (54 mg, 64%) as a yellow solid: mp 104.1-105.2 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.35 (t, *J* = 7.2 Hz, 3H, CH_2CH_3), 4.15 (s, 3H, OCH_3), 4.36 (q, *J* = 7.1 Hz, 2H, CH_2CH_3), 7.05 (d, *J* = 8.0 Hz, 1H, *H*-6), 7.39 (d, *J* = 8.0 Hz, 1H, *H*-7), 7.43 (m, 5H, *aryl*), 7.48 (d, *J* = 1.2 Hz, 1H, *H*-3), 8.22 (d, *J* = 1.2 Hz, 1H, *H*-1), 9.49 (s, 1H, O*H*); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 14.24 (CH_2CH_3), 56.42 (OCH_3), 61.16 (CH_2CH_3), 102.94 (*C*-3), 112.65 (*C*-6), 116.81 (*C*-4a), 123.30 (*C*-1), 127.09 (*C*-7), 127.57 (*C*-2), 128.29 (*C*-2'), 129.80 (*C*-4'), 130.20 (*C*-3'), 132.98 (*C*-1'), 133.52 (*C*-8), 140.43 (*C*-8a), 153.98 (*C*-5), 156.41 (*C*-4), 166.43 (*C*OO); *m/z* (CI^+) 323 (M^+, 100), 309 (6); HRMS (CI^+) calcd for $C_{20}H_{19}O_4$: 323.1283, found 323.1277.

Ethyl 5-hydroxy-4-methoxy-8-naphthyl-2-naphthalenecarboxylate (100)

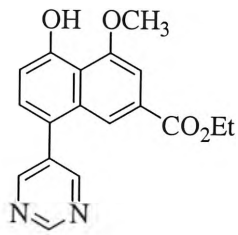


(**100**)

To a solution of (**72**) (45 mg, 109 μmol) in DCM (20 mL) was added dropwise boron trifluoride diethyl etherate (125μL, 1.06 mmol) under N_2, at 0°C. The resulting solution was allowed to stir at room temperature for 24 hours, followed by the addition of methanol (5 mL). The solvent was removed under reduced pressure, and the residue extracted with chloroform (4 x 20 mL) and water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (20% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**100**) (36 mg, 89%) as a yellow solid: mp 174.0-176.8 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.21 (t, *J* = 7.2 Hz, 3H, CH_2CH_3), 4.17 (s, 3H, OCH_3), 4.23 (q, *J* = 7.1 Hz, 2H, CH_2CH_3), 7.12 (d, *J* = 8.0 Hz, 1H, *H*-6), 7.31 (d, *J* = 6.8 Hz, 1H, *naphthyl*), 7.39 (d, *J* = 8.0 Hz, 1H, *H*-7), 7.45 (d, *J* = 1.6 Hz, 1H, *H*-3),

7.45 (m, 1H, *naphthyl*), 7.59 (t, *J* = 6.8 Hz, 1H, *naphthyl*), 7.73 (q, *J* = 6.4 Hz, 2H, *naphthyl*), 7.81 (s, 1H, *H*-1), 7.95 (d, *J* = 8.4 Hz, 2H, *naphthyl*), 9.55 (s, 1H, O*H*); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.09 (CH_2CH_3), 56.47 (OCH_3), 61.03 (CH_2CH_3), 102.95 (*C*-3), 112.53 (*C*-6), 123.64 (*C*1), 125.30 (*naphthyl*), 125.72, 125.95, 126.25, 127.65, 127.95, 128.17, 128.34, 130.77, 130.82, 133.12, 133.69, 134.82, 135.11, 137.87, 154.22 (*C*-5), 156.43 (*C*-4), 166.29 (*C*OO); *m/z* (CI^+) 373 (*M*+1, 26), 263 (100); HRMS (CI^+) calcd for $C_{24}H_{21}O_4$: 373.1440, found 373.1441.

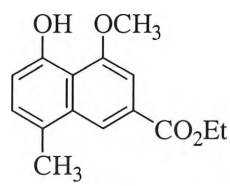
Ethyl 5-hydroxy-4-methoxy-8-(5'-pyrimidyl)-2-naphthalenecarboxylate (101)



(**101**)

To a solution of (**83**) (50 mg, 136 μmol) in DCM (10 mL) was added dropwise boron trifluoride diethyl etherate (35 μL, 276 μmol) under N_2, at 0°C. The resulting solution was allowed to stir at room temperature for 24 hours, followed by the addition of methanol (5 mL). The solvent was removed under reduced pressure, and the residue extracted with chloroform (4 x 20 mL) and water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (20% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**101**) (25 mg, 56%) as a pale yellow solid: decomposes 219 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.36 (t, *J* = 7.0 Hz, 3H, CH_2CH_3), 4.17 (s, 3H, OCH_3), 4.37 (q, *J* = 7.1 Hz, 2H, CH_2CH_3), 7.08 (d, *J* = 8.4 Hz, 1H, *H*-6), 7.37 (d, *J* = 8.0 Hz, 1H, *H*-7), 7.45 (d, *J* = 0.8 Hz, 1H, *H*-3), 8.05 (d, *J* = 1.6 Hz, 1H, *H*-1), 8.83 (s, 2H, *H*-4', *H*-6'), 9.28 (s, 1H, *H*-2'), 9.61 (s, 1H, O*H*); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 14.27 (CH_2CH_3), 56.66 (OCH_3), 61.51 (CH_2CH_3), 103.75 (*C*-3), 112.81 (*C*-6), 117.06 (*C*-8a), 121.33 (*C*-1), 124.78 (*C*-4a), 128.85 (*C*-2), 130.74 (*C*-7, *C*-1'), 133.42 (*C*-8), 155.61 (*C*-5), 156.72 (*C*-4), 157.42 (*C*-2', *C*-4'), 165.97 (*C*OO); *m/z* (CI^+) 325 (M^+, 100), 311 (5); HRMS (CI^+) calcd for $C_{18}H_{17}N_2O_4$: 325.1188, found 325.1184.

Ethyl 5-hydroxy-4-methoxy-8-methyl-2-naphthalenecarboxylate (102)



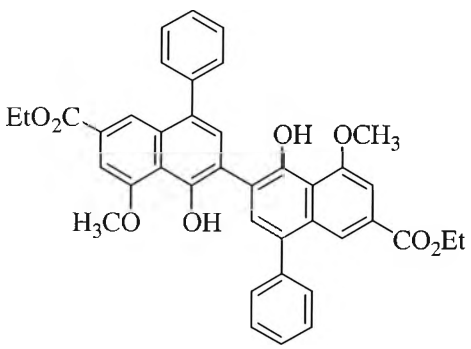
(**102**)

To a solution of (**79**) (190 mg, 628 μmol) in DCM (30 mL) was added dropwise boron trifluoride diethyl etherate (300 μL, 2.54 mmol) under N_2, at 0°C. The resulting solution was allowed to stir at room temperature for 24 hours, followed by the addition of methanol (10 mL). The solvent was removed under reduced

pressure, and the residue extracted with chloroform (4 x 30 mL) and water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (10% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**102**) (89 mg, 55%) as a pale green solid: mp 71.6-73.8 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.45 (t, *J* = 7.0 Hz, 3H, CH_2CH_3), 2.62 (s, 3H, Ar-CH_3), 4.12 (s, 3H, OCH_3), 4.45 (q, *J* = 7.2 Hz, 2H, CH_2CH_3), 7.90 (d, *J* = 7.6 Hz, 1H, *H*-6), 7.26 (d, *J* = 8.0 Hz, 1H, *H*-7), 7.38 (d, *J* = 1.2 Hz, 1H, *H*-3), 8.33 (d, *J* = 1.6 Hz, 1H, *H*-1), 9.25 (s, 1H, *OH*); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 14.40 (CH_2CH_3), 19.38 (Ar-CH_3), 56.35 (OCH_3), 61.29 (CH_2CH_3), 102.92 (*C*-3), 112.51 (*C*-6), 116.97 (*C*-4a), 121.47 (*C*-1), 126.39 (*C*-2), 127.27 (*C*-8a), 129.20 (*C*-7), 134.28 (*C*-8), 152.85 (*C*-5), 156.63 (*C*-4), 166.63 (*COO*); *m/z* (CI^+) 261 (*M*+1, 100), 247 (33); HRMS (CI^+) calcd for $C_{15}H_{17}O_4$: 261.1127, found 261.1122. (MA found: C, 69.42; H, 6.47; O, 24.41; $C_{15}H_{16}O_4$ requires C, 69.22; H, 6.20; O, 24.59).

Diethyl 5,5'-dihydroxy-4,4'-dimethoxy-8,8'-diphenyl-(6,6'-binaphthyl)-2,2'-dicarboxylate (103)

To a solution of (**91**) (40 mg, 55 μmol) in DCM (15 mL) was added dropwise boron trifluoride diethyl etherate (60 μL, 507 mmol) under N_2, at 0°C. The resulting solution was allowed to stir at room temperature for 24 hours, followed by the addition of methanol (5 mL). The solvent was removed under reduced pressure, and the residue extracted with chloroform (4 x 20 mL) and water.

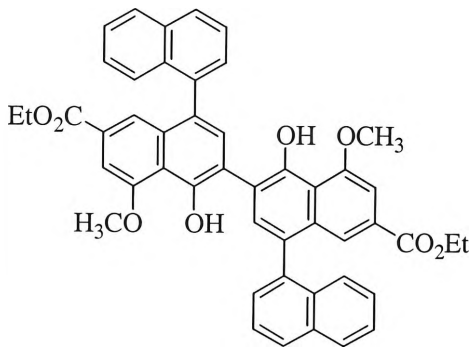


(**103**)

The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (20% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**103**) (24 mg, 68%) as a green solid: mp 249.8-252.2 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.37 (t, *J* = 7.2 Hz, 6H, CH_2CH_3, $CH'_2CH'_3$), 4.14 (s, 6H, OCH_3, OCH'_3), 4.37 (q, *J* = 7.2 Hz, 4H, CH_2CH_3, $CH'_2CH'_3$), 7.43 (d, *J* = 1.2 Hz, 2H, *H*-3, *H*-3'), 7.46 (m, 10H, *aryl''*, *aryl'''*), 7.59 (s, 2H, *H*-7, *H*-7'), 8.35 (d, *J* = 1.6 Hz, 2H, *H*-1, *H*-1'), 9.99 (s, 2H, *OH*, *OH'*); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 14.27 (CH_2CH_3, $C'H_2C'H_3$), 56.54 (OCH_3, $OC'H_3$), 61.20 (CH_2CH_3, $C'H_2C'H_3$), 103.38 (*C*-3, *C*-3'), 117.12 (*C*-6, *C*-6'), 122.11 (*C*-4a, *C*-4a'), 123.37 (*C*-1, *C*-1'), 127.11 (*C*-4'', *C*-4'''), 127.58 (*C*-2, *C*-2'), 128.29 (*C*-2'', *C*-6'', *C*-2''', *C*-6'''),

130.35 (*C*-3'', *C*-5'', *C*-3''', *C*-5'''), 132.61 (*C*-1'', *C*-1'''), 132.66 (*C*-7, *C*-7'), 133.29 (*C*-8, *C*-8'), 140.31 (*C*-8a, 8a'), 150.92 (*C*-5, *C*-5'), 156.71 (*C*-4, *C*-4'), 166.52 (*C*OO, *C*'OO); *m/z* (ES^+) 643 (*M*+1, 8), 338 (89), 225 (23); HRMS (ES^+) calcd for $C_{40}H_{35}O_8$: 643.2332, found 643.2323. (MA found: C, 74.72; H, 5.52; O, 19.76; $C_{40}H_{34}O_8$ requires C, 74.75; H, 5.33; O, 19.92).

Diethyl 5,5'-dihydroxy-4,4'-dimethoxy-8,8'-(1'',1'''-dinaphthyl)-(6,6'-binaphthyl)-2,2'-dicarboxylate (104)

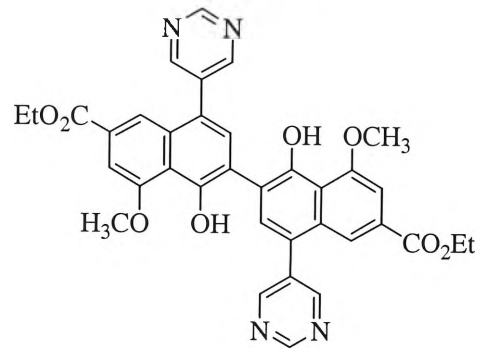


(**104**)

To a solution of (**96**) (180 mg, 218 μmol) in DCM (30 mL) was added dropwise boron trifluoride diethyl etherate (205μL, 1.73mmol) under N_2, at 0°C. The resulting solution was allowed to stir at room temperature for 24 hours, followed by the addition of methanol (10 mL). The solvent was removed under reduced pressure, and the residue extracted with chloroform (4 x 20 mL) and water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (25% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**104**) (110 mg, 68%) as a yellow solid; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.22 (t, *J* = 7.1 Hz, 6H, CH_2CH_3, $CH'_2CH'_3$), 4.15 (s, 6H, OCH_3, OCH'_3), 4.23 (q, *J* = 7.1 Hz, 4H, CH_2CH_3, $CH'_2CH'_3$), 7.31 (m, 2H, *naphthyl''*, *naphthyl'''*), 7.39 (d, *J* = 1.2 Hz, 2H, *H*-3, *H*-3'), 7.54 (m, 8H, *naphthyl''*, *naphthyl'''*), 7.66 (s, 2H, *H*-7, *H*-7'), 7.82 (d, *J* = 1.5 Hz, 2H, *H*-1, *H*-1'), 7.91 (d, *J* = 7.8 Hz, 4H, *naphthyl''*, *naphthyl'''*), 10.05 (s, 2H, O*H*, O*H*'); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.12 (CH_2CH_3, $C'H_2C'H_3$), 56.52 (OCH_3, $OC'H_3$), 61.05 (CH_2CH_3, $C'H_2C'H_3$), 103.25 (*C*-3, *C*-3'), 117.00 (*C*-6, *C*-6'), 121.94, 123.65 (*C*-1, *C*-1'), 125.31 (*naphthyl''*, *naphthyl'''*), 125.67, 126.00, 126.43, 127.54, 127.93, 128.12, 128.50, 130.34, 133.13, 133.60, 133.67, 134.50 (*C*-7, *C*-7'), 137.73, 151.16 (*C*-5, *C*-5'), 156.64 (*C*-4, *C*-4'), 166.36 (*C*OO, *C*'OO); *m/z* (ES^+) 743 (*M*+1, 100), 675 (63); HRMS (ES^+) calcd for $C_{48}H_{39}O_8$: 743.2645, found 743.2654. (MA found: C, 77.68; H, 5.43; O, 16.89; $C_{48}H_{39}O_8$ requires C, 77.61; H, 5.16; O, 17.23).

Diethyl 5,5'-dihydroxy-4,4'-dimethoxy-8,8'-(5'',5'''-dipyrimidyl)-(6,6'-binaphthyl)-2,2'-dicarboxylate (105)

To a solution (**97**) (150 mg, 205 μmol) in DCM (30 mL) was added dropwise boron trifluoride diethyl etherate (195 μL, 1.65 mmol) under N_2, at 0°C. The resulting solution was allowed to stir at room temperature for 24 hours, followed by the addition of methanol (10 mL). The solvent was removed under reduced pressure, and the residue extracted with chloroform (4 x 20 mL) and water.

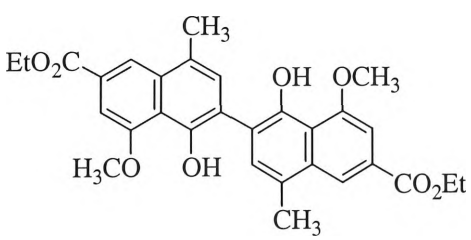


(**105**)

The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (80% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**105**) (87 mg, 65%) as a pale green solid: mp 267.4-268.8 °C; ^{1}H NMR ($CDCl_3$, 300 MHz) δ 1.38 (t, *J* = 7.2 Hz, 6H, CH_2CH_3, $CH'_2CH'_3$), 4.18 (s, 6H, OCH_3, OCH'_3), 4.40 (q, *J* = 7.1 Hz, 4H, CH_2CH_3, $CH'_2CH'_3$), 7.51 (d, *J* = 1.2 Hz, *H*-3, *H*-3'), 7.60 (s, 2H, *H*-7, *H*-7'), 8.11 (d, *J* = 1.5 Hz, 2H, *H*-1, *H*-1'), 8.91 (s, 4H, *H*-4'', *H*-4''', *H*-6'', *H*-6'''), 9.28 (s, 2H, *H*-2'', *H*-2'''), 10.13 (s, 2H, O*H*, O*H*;); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 14.31 (CH_2CH_3, $C'H_2C'H_3$), 56.79 (OCH_3, $OC'H_3$), 61.58 (CH_2CH_3, $C'H_2C'H_3$), 104.27 (*C*-3, *C*-3'), 117.38 (*C*-8a, *C*-8a'), 121.31 (*C*-1, *C*-1'), 121.34 (*C*-4a, *C*-4a'), 124.55 (*C*-8, *C*-8'), 128.98 (*C*-6, *C*-6'), 133.14 (*C*-7, *C*-7'), 133.37 (*C*-2, *C*-2'), 134.08 (*C*-5'', *C*-5'''), 152.46 (*C*-5, *C*-5'), 156.93 (*C*-2'', *C*-2'''), 157.52 (*C*-4, *C*-4', *C*-4'', *C*-4''', *C*-5'', *C*-5'''), 165.95 (*C*OO, *C'*OO); *m/z* (ES^+) 647 (*M*+1, 33), 633 (4), 324 (100); HRMS (ES^+) calcd for $C_{36}H_{31}N_4O_8$: 647.2142, found 647.2132.

Diethyl 5,5'-dihydroxy-4,4'-dimethoxy-8,8'-dimethyl-(6,6'-binaphthyl)-2,2'-dicarboxylate (106)

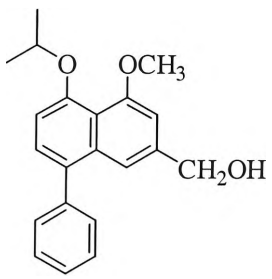
To a solution of (**98**) (85 mg, 141 μmol) in DCM (15 mL) was added dropwise boron trifluoride diethyl etherate (135 μL, 1.14 mmol) under N_2, at 0°C. The resulting solution was allowed to stir at room temperature for 24 hours, followed by the addition of methanol (5 mL). The solvent was



(**106**)

removed under reduced pressure, and the residue extracted with chloroform (4 x 20 mL) and water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (20% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**106**) (52 mg, 71%) as a yellow solid: mp 247.6-249.7 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.47 (t, J = 7.2 Hz, 6H, CH_2CH_3, $CH'_2CH'_3$), 2.67 (s, 6H, Ar-CH_3, Ar'-CH'_3), 4.11 (s, 6H, OCH_3, OCH'_3), 4.47 (q, J = 7.3 Hz, 4H, CH_2CH_3, $CH'_2CH'_3$), 7.39 (s, 2H, *H*-3, *H*-3'), 7.43 (s, 2H, *H*-7, *H*-7'), 8.38 (s, 2H, *H*-1, *H*-1'), 9.72 (s, 2H, O*H*, O*H'*); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 14.43 (CH_2CH_3, $C'H_2C'H_3$), 19.42 (Ar-CH_3, Ar'-$C'H_3$), 56.41 (OCH_3, $OC'H_3$), 61.30 (CH_2CH_3, $C'H_2C'H_3$), 103.35 (*C*-3, *C*-3'), 117.24 (*C*-4a, *C*-4a'), 121.51 (*C*-1, *C*-1'), 122.22 (*C*-8a, *C*-8a'), 126.04 (*C*-6, *C*-6'), 127.24 (*C*-2, *C*-2'), 131.89 (*C*-7, *C*-7'), 134.09 (*C*-8, *C*-8'), 149.65 (*C*-5, *C*-5'), 156.85 (*C*-4, *C*-4'), 166.69 (*C*OO, *C'*OO); *m/z* (CI^+) 519 (*M*+1, 100), 505 (8); HRMS (CI^+) calcd for $C_{30}H_{31}O_8$: 519.2019, found 519.2018. (MA found: C, 69.57; H, 5.86; O, 24.57; $C_{30}H_{31}O_8$ requires C, 69.49; H, 5.83; O, 24.68).

5-Isopropoxy-4-methoxy-8-phenylnaphthalene-2-methanol (107)

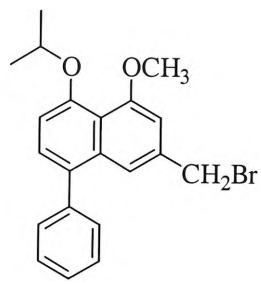


(**107**)

To a solution of (**68**) (170 mg, 466 μmol) in dry THF (60 mL) was added a suspension of lithium aluminium hydride (18 mg, 474 μmol) in THF (5 mL) under N_2, at 0 °C. The resulting solution was allowed to stir at room temperature for 1 hour followed by the addition of 1M HCl (10 mL). The solvent was removed under reduced pressure, and the residue extracted with DCM (4 x 20 mL) and water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before chromatography with a short silica plug (30% ethyl acetate:hexane). Recrystallisation from isopropanol afforded (**107**) (137 mg, 91%) as yellow solid: mp 112.2-114.1 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.44 (d, J = 6.0 Hz, 6H, $CH(CH_3)_2$), 3.98 (s, 3H, OCH_3), 4.58 (p, J = 6.1 Hz, 1H, $CH(CH_3)_2$), 4.67 (s, 2H, CH_2OH), 6.89 (s, 1H, *H*-3), 6.97 (d, J = 8.0 Hz, 1H, *H*-6), 7.28 (d, J = 8.0 Hz, 1H, *H*-7), 7.34 (d, J = 1.6 Hz, 1H, *H*-1), 7.43 (m, 5H, *aryl*); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 22.15 ($CH(CH_3)_2$), 56.43 (OCH_3), 65.71 (CH_2OH), 73.21 ($CH(CH_3)_2$), 105.51 (*C*-3), 112.71 (*C*-6), 116.70 (*C*-7), 119.09 (*C*-4a), 126.88 (*C*-4'), 127.99 (*C*-1), 128.20 (*C*-2', 6'), 130.23 (*C*-3', 5'), 133.72 (*C*-1'), 135.37 (*C*-8), 138.64 (*C*-2), 141.44 (*C*-8a),

154.51 (*C*-5), 157.75 (*C*-4); *m/z* (CI^+) 323 (*M*+1, 100), 305 (79), 281 (49); HRMS (CI^+) calcd for $C_{21}H_{23}O_3$: 323.1647, found 323.1641.

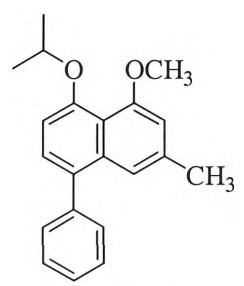
5-Isopropoxy-4-methoxy-2-methylbromide-8-phenylnaphthalene (108)



(108)

To a solution of (**107**) (140 mg, 434 μmol) in dry distilled ether (30 mL) was added triphenylphosphine (115 mg, 438 μmol) and carbon tetrabromide (145 mg, 437 μmol) under N_2, at room temperature. The resulting solution was allowed to stir for 4 hours. The solvent was removed under reduced pressure, and the extracted with chloroform (4 x 30 mL) and water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (20% ethyl acetate:hexane). Recrystallisation in isopropanol yielded (**108**) (158 mg, 94%) as a yellow solid: mp 92.9-94.8 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.43 (d, *J* = 6.0 Hz, 6H, $CH(CH_3)_2$), 3.99 (s, 3H, OCH_3), 4.50 (s, 2H, CH_2Br), 4.58 (p, *J* = 6.1 Hz, 1H, $CH(CH_3)_2$), 6.86 (d, *J* = 1.6 Hz, 1H, *H*-3), 6.99 (d, *J* = 8.0 Hz, 1H, *H*-6), 7.29 (d, *J* = 8.0 Hz, 1H, *H*-7), 7.39 (d, *J* = 1.6 Hz, 1H, *H*-1), 7.44 (m, 5H, *aryl*); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 22.14 ($CH(CH_3)_2$), 34.64 (CH_2Br), 56.41 (OCH_3), 73.24 ($CH(CH_3)_2$), 106.84 (*C*-3), 113.36 (*C*-6), 119.03 (*C*-7), 126.98 (*C*-4'), 128.26 (*C*-2', *C*-6'), 128.30 (*C*-1), 128.40 (*C*-4a), 130.40 (*C*-1'), 130.24 (*C*-3', *C*-5'), 133.84 (*C*-2), 135.26 (*C*-8), 141.16 (*C*-8a), 154.53 (*C*-5), 157.90 (*C*-4); *m/z* (CI^+) 387 (*M*+1, Br^{81}, 98), 385 (*M*+1, Br^{79}, 100), 345 (Br^{81}, 52), 343 (Br^{79}, 56), 305 (52), 265 (9); HRMS (CI^+) calcd for $C_{21}H_{22}BrO_2$: 385.0803, found 385.0799.

5-Isopropoxy-4-methoxy-2-methyl-8-phenylnaphthalene (109)

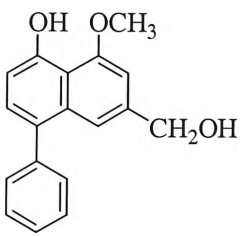


(109)

To a solution of (**108**) (25 mg, 65 μmol) in dry DCM (15 mL) was added a 1.0 M solution of L-selectride in THF (15 μL, 70 μmol) under N_2, at 0 °C. The resulting solution was allowed to stir at room temperature for 2 hours followed by the addition of a second equivalent of L-selectride (15 μL, 70 μmol) under N_2, at 0 °C. The solution was allowed to stir for 1 hour, followed by the addition of ethanol (5 mL). The solvent was removed under reduced pressure, and the residue was extracted with chloroform (4 x 20 mL). The organic phase was

dried ($MgSO_4$), and concentrated *in vacuo*, before being subjected to flash column chromatography (10% ethyl acetate:hexane). Recrystallisation from isopropanol yielded (**109**) (19 mg, 96%) as a yellow solid: mp 70.9-73.7 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 1.44 (d, *J* = 6.0 Hz, 6H, CH(C*H*$_3$)$_2$), 2.36 (s, 3H, C*H*$_3$), 3.96 (s, 3H, OC*H*$_3$), 4.58 (p, *J* = 6.0 Hz, 1H, C*H*(CH$_3$)$_2$), 6.70 (d, *J* = 1.2 Hz, 1H, *H*-3), 6.91 (d, *J* = 8.0 Hz, 1H, *H*-6), 7.19 (d, *J* = 1.2 Hz, 1H, *H*-1), 7.24 (d, *J* = 8.0 Hz, 1H, *H*-7), 7.45 (m, 5H, *aryl*); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 22.02 (*C*H$_3$), 22.17 (CH(*C*H$_3$)$_2$), 56.57 (O*C*H$_3$), 73.04 (*C*H(CH$_3$)$_2$), 108.99 (*C*-3), 111.78 (*C*-6), 117.86 (*C*-2), 118.33 (*C*-7), 126.69 (*C*-4'), 127.69 (*C*-1), 128.10 (*C*-2', 6'), 130.27 (*C*-3', 5'), 133.06 (*C*-1'), 135.62 (*C*-4a), 135.87 (*C*-8), 141.78 (*C*-8a), 154.45 (*C*-5), 157.07 (*C*-4); *m/z* (CI$^+$) 307 (*M*+1, 100), 265 (38); HRMS (CI$^+$) calcd for $C_{21}H_{23}O_2$: 307.1698, found 307.1698.

5-Hydroxy-4-methoxy-8-phenylnaphthalene-2-methanol (110)



(110)

To a solution of (**107**) (100 mg, 310 μmol) in dry THF (40 mL) was added a suspension of lithium aluminium hydride (12 mg, 316 μmol) in THF (5 mL) under N_2, at 0 °C. The resulting mixture was allowed to stir at room temperature for 1 hour followed by the addition of 1M HCl (10 mL). The solvent was removed under reduced pressure, and the residue extracted with DCM (4 x 20 mL) and water. The organic phase was dried ($MgSO_4$), and concentrated *in vacuo*, before chromatography with a short silica plug (50% ethyl acetate:hexane). Recrystallisation from isopropanol afforded (**110**) (46 mg, 53%) as yellow solid: mp 152.0-154.4 °C; ^{1}H NMR ($CDCl_3$, 400 MHz) δ 4.08 (s, 2H, C*H*$_2$OH), 4.67 (s, 3H, OC*H*$_3$), 6.87 (s, 1H, *H*-3), 6.93 (d, *J* = 7.6 Hz, 1H, *H*-6), 7.30 (d, *J* = 8.0 Hz, 1H, *H*-7), 7.39 (s, 1H, *H*-1), 7.43 (m, 5H, *aryl*), 9.48 (s, 1H, O*H*); ^{13}C NMR ($CDCl_3$, 100 MHz) δ 56.23 (O*C*H$_3$), 65.56 (*C*H$_2$OH), 103.22 (*C*-3), 110.25 (*C*-6), 114.47 (*C*-4a), 117.530 (*C*-7), 126.84 (*C*-1), 128.25 (*C*-2', *C*-6'), 129.42 (*C*-4'), 130.21 (*C*-3', *C*-5'), 131.36 (*C*-1'), 134.35 (*C*-8), 138.43 (*C*-2), 141.12 (*C*-8a), 154.05 (*C*-5), 156.77 (*C*-4); *m/z* (CI$^+$) 281 (*M*+1, 100), 263 (68); HRMS (CI$^+$) calcd for $C_{18}H_{17}O_3$: 281.117770, found 281.118046.

6.2 Computer modelling

Atomic coordinates for all three-dimensional x-ray crystal structures were obtained from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank.[265,320] A total of 23 x-ray crystallographic files were used, and some statistics relevant to the structures are tabulated in Appendix 7.3. Three additional crystal structures were used for only selected purposes, such as pharmacophore validation (Chapter 4), and these are tabulated in Appendix 7.3.1.

All computer modelling was performed on Silicon Graphics O2 workstations (Irix Release 6.5), with software as mentioned in the text.

6.2.1 Modelling of protein conformational changes

The superimposition subset for the conformational studies of reverse transcriptase was derived using the ProFlex program to calculate DMs and DDMs of different x-ray crystal structures of RT.[147,151] The final superimposition subset is listed in Table 3.1 of Section 3.2.2. Superimposition was performed in the InsightII® program,[152] by overlaying the second unliganded structure (1hmv), both DNA-bound structures (1rtd and 2hmi) and all of the inhibitor-bound structures onto the chosen unliganded structure, 1dlo, with this subset. RMSD values were obtained from InsightII for the individual superimpositions and are listed in Table 3.2 of Section 3.2.2. The distance between corresponding atoms of equivalent residues in each individual crystal structure and the unliganded 1dlo was measured once the crystal structure had been superimposed onto 1dlo. Only the atoms of those residues in the active sites and motifs listed in Appendix 7.6, Subset definitions, were analysed. These distances were tabulated and graphed in Excel®.[153]

6.2.2 Ligand-based pharmacophore generation with Catalyst

The ligand-based pharmacophore was generated using the Catalyst® software program (Molecular Simulations Inc.).[291,307]

The inhibitor heteroatom coordinates of all selected liganded PDB files were saved in Sybyl mol2 file format after the correct atom types were applied in InsightII® (Molecular Simulations Inc.).[152] These were imported into Catalyst in their original crystal conformations. Conformational models were generated in the Best-quality

mode, with the default energy range of 20 kcal.mol^{-1} from the minimum, based on the CHARMm forcefield.

Catalyst pharmacophores were generated using hydrogen bond acceptors, hydrogen bond donors, hydrophobic aliphatic, positive ionisable, and ring aromatic groups as the five possible feature types. All standard HypoGen parameters[297,298] were used with the exception of the *Minimum Total Features* parameter and the *forceAbsoluteSterochemistry* option. The *Minimum Total Features* parameter was set at varying values as discussed in the text (see Section 4.2.3 Hypothesis generation) and increased until such a value where the hypothesis generation failed due to unfavourable cost values. The *forceAbsoluteSterochemistry* option was turned off by adding 'catHypo.forceAbsoluteStereochemistry=1' to the .Catalyst file.

Database searching was performed with the Best Flexible Searching Tool, with the *Maximum Search Hits* option set at 10 000. The results from database searches can be saved in a spreadsheet format, or as a database, to be searched by subsequent pharmacophores.

6.2.3 Structure-based pharmacophore generation with Unity 3D

A set of residues within a 25 Å radius of the inhibitor was defined for each of the nineteen liganded crystal structures of Appendix 7.3, except 3hvt, and all other protein residues were deleted. Using the Biopolymer module of Sybyl, hydrogens and charges (Kollman_All charge set) were added before the hydrogens were minimised. Minimisation calculations were carried out with the Powell method, using the Tripos forcefield, and were terminated once the RMS reached 0.05 kcal.mol^{-1}.Å or after 500 iterations.

Binding pocket residues were identified using the SiteID® module of Sybyl.[309] The entire structure of each of the truncated crystal structures was selected for the initial solvation in SiteID. All standard parameters were used with the exception of the *Minimum Pocket Concavity* and the *van der Waals bump scaling* factor. Solvation of each crystal structure was performed with the *Minimum Pocket Concavity* and the *van der Waals bump scaling* factor set at 2 and 0.8 Å, respectively, and 10 and 0.7 Å, respectively.

The 'largest' and 'smallest' modified binding pockets, based on the position of the backbone and sidechain, were generated in InsightII by initially superimposing all the RT-inhibitor crystal structures onto the unliganded 1dlo structure (Section 6.2.1). All residues, excepting those identified as binding pocket residues by SiteID, were deleted. The coordinates of the remaining liganded crystal structures were transformed, so that they were relative to each other, and were saved in PDB format. The average point for the backbone and sidechain of each residue of each individual binding pocket was calculated and its distance from a central inhibitor reference point was determined. The crystal structure with the minimum and maximum distance values for the backbone and sidechain of each residue was noted in the Table 6.1, and the complete coordinates for this residue from this crystal structure were then used for the modified pockets. The file was saved in PDB format for use with Unity 3D.

Residue Details		Backbone		Sidechain	
		minimum	maximum	minimum	maximum
Trp	88	hni	rti	hni	rev
Glu	89	rth	rt2	vru	rev
Val	90	bqm	rt3	rev	rt2
Gln	91	klm	rt3	klm	rev
Leu	92	rt7	uwb	rt4	uwb
Gly	93	rt5	hni	rt5	hni
Ile	94	rti	hni	rt5	hni
Pro	95	rti	hni	rt4	hni
His	96	klm	uwb	vru	tvr
Pro	97	vru	rt6	hni	rt1
Gly	99	vru	rt6	vru	rt6
Leu	100	rev	rt6	vru	hni
Lys	101	bqn	klm	rev	tvr
Lys	102	tvr	rti	tvr	rti
Lys	103	bqm	rt4	hni	klm
Val	106	tvr	vru	rt1	vru
Gln	161	rt7	uwb	rt1	uwb
Ser	162	rt7	bqm	rti	tvr
Met	164	klm	tvr	klm	bqm
Thr	165	hni	rt1	hni	rt4
Leu	168	rt3	rth	rt3	rev
Arg	172	rt3	rth	rt3	uwb [§]
Val	179	hni	rth	hni	rt6
Ile	180	hni	rt3	rti	bqn
Tyr	181	rti	tvr [†]	rti	hnv
Gln	182	rt3	hnv	rti	tvr
Tyr	183	rt3	bqm	rt3	bqn
Met	184	rt3	bqm	rt3	tvr
Asp	186	rti	hnv	rti	uwb

Leu	187	klm	hnv	rti	tvr
Tyr	188	klm	rev	rti	hnv
Val	189	rt3	vru	rt3	bqn
Gly	190	uwb	rth	uwb	rth
Glu	224	rt4 †	hni	rt6 †	rt3
Pro	225	bqn	klm	rt4	tvr
Pro	226	rt6	hni	rt6	uwb
Phe	227	klm	bqm	rt2	bqn
Trp	229	klm	tvr	klm	uwb
Met	230	rth	tvr	klm	tvr
Tyr	232	hni	tvr	hni	rt1
Leu	234	rt3	bqn	rt3	bqn
His	235	hnv	klm	uwb	klm
Pro	236	uwb	klm	uwb	klm
Tyr	318	uwb	rt6	uwb	klm
Ile	380	vru	tvr	bqn	tvr
Val	381	vru	tvr	klm	tvr
Ile	382	vru	hnv	vru	rt7
Pro	25	bqn	rev	bqn	rt1
Leu	26	rth	uwb	rth	uwb
Ile	31	bqn	rt3	tvr	rt3
Thr	131	rt4	uwb	rth	hni
Ile	132	rt7	hni	rt7	bqn
Pro	133	rt5	hni	rt5	hni
Ser	134	rt5	rt3	rt7	hni
Ile	135	rti	tvr	rti	rt3
Asn	136	rt4	uwb	rti	uwb
Asn	137	rt4	uwb	rt4	uwb
Glu	138	rt4	hni	klm	rti
Thr	139	rth	hni	rti	hni
Pro	140	rth	vru	tvr	vru
Gly	141	tvr	rt2	tvr	rt2
Ile	142	rth	hni	rth	uwb
Arg	143	rt2	tvr	hnv	bqm

Table 6.1 Lists of the crystal structures from which each complete residue of the modified binding pockets was derived.
§ 1uwb has a lysine sidechain instead of arginine
† The sidechain of the residue whose distance showed maximal or minimal distance from the inhibitor reference point was missing. The crystal structure with the next greatest maximum or minimum distance was used

Structural queries were generated in Unity 3D[280] by using the Constraints and Select Donor/Acceptor tools. The Constraints tool was used to define excluded volumes, and standard parameters were selected, with the exception of the van der Waals scaling factor that was used with the default of 1.0 and set at 0.5. The Select Donor/Acceptor tool was used to perform an automated search for all donor and acceptor sites on the protein. Only those sites that projected the potential feature of the ligand into the binding pocket were chosen. All of the donor and acceptor sites on the protein chosen

as features in the structural query were selected for Unity's Partial Match Constraint Dialog by typing 'define unity_feature partial_match_constraint m2 %unity_feature_names(m2 macro) |' at the Sybyl prompt. Structural queries can be saved in the Sybyl mol2 file format, and provide Cartesian coordinates and spatial information of the features of the query that can be used in other programs such as Catalyst (see Section 4.3.6 Structural queries in Catalyst).

A Flex Query search using the Partial Match Constraint Dialog was performed in Unity 3D[280] using standard parameters with the exception of the minimum and maximum features to be accepted as a hit. These were set at varying valuables as noted in the text (see Section 4.3.4 Database searching with Unity queries). The time spent attempting to fit each compound to the structural query was set at 90 seconds. Search results were saved in a database format.

Unity 3D structural queries were converted to Catalyst pharmacophores by saving the queries in the Sybyl mol2 file format. The coordinates of the features were expressed in angstroms and needed to be multiplied by 100 to convert them to picometres, the units used for the Hypoedit utility (see Appendix 7.14). The Hypoedit file was modified to include the features and excluded volumes data from the structural query, and a pharmacophore was generated by typing 'hypoedit' at the unix prompt and pasting the modified Hypoedit file (see Appendix 7.14.1 Hypoedit file of activity 'weighted' structural query).

References

1. Tang, H. L., Kuhen, K. L. & Wong-Staal, F. Lentivirus replication and regulation. *Ann. Rev. Gen.***33**, 133-170 (1999).

2. Pavlakis, G. N. "The molecular biology of human immunodeficiency virus type 1." in *AIDS: Biology, Diagnosis, Treatment and Prevention* (eds. DeVita, J. V. T., Hellman, S. & Rosenberg, S. A.) (Lippincott-Raven, Philadelphia, 1996).

3. Turner, B. G. & Summers, M. F. Structural biology of HIV. *J. Mol. Biol.* **285**, 1-32 (1999).

4. Wilk, T. & Fuller, S. D. Towards the structure of the human immunodeficiency virus: divide and conquer? *Curr. Opin. Struct. Biol.* **9**, 231-243 (1999).

5. Hottiger, M. & Hubscher, U. Human immunodeficiency virus type 1 reverse transcriptase. *Biol. Chem. Hoppe-Seyler* **377**, 97-120 (1996).

6. Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M. & Markowitz, M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* **373**, 123-126 (1995).

7. Wei, X. P., Ghosh, S. K., Taylor, M. E., Johnson, V. A., Emini, E. A., Deutsch, P., Lifson, J. D., Bonhoeffer, S., Nowak, M. A., Hahn, B. H., Saag, M. S. & Shaw, G. M. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373**, 117-122 (1995).

8. Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. HIV-1 dynamics *in vivo* - Virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**, 1582-1586 (1996).

9. Stewart, G. (ed.) *Managing HIV* (Australasian Medical Publishing Company Limited, Sydney, 1997).

10. Wodarz, D. & Nowak, M. A. Evolutionary dynamics of HIV-induced subversion of the immune response. *Immun. Rev.* **168**, 75-89 (1999).

11. Ho, W., Kukla, M. J., Breslin, H. J., Ludovici, D. W., Grous, P. P., Diamond, C. J., Miranda, M., Rodgers, J. D., Ho, C. Y., Declercq, E., Pauwels, R., Andries, K., Janssen, M. A. C. & Janssen, P. A. J. Synthesis and anti-HIV-1 activity of 4,5,6,7-tetrahydro-5-methylimidazo-[4,5,1-*jk*][1,4]benzodiazepin-2(1*H*)-one (TIBO) derivatives .4. *J. Med. Chem.* **38**, 794-802 (1995).

12. Chun, T. W., Stuyver, L., Mizell, S. B., Ehler, L. A., Mican, J. A. M., Baseler, M., Lloyd, A. L., Nowak, M. A. & Fauci, A. S. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proceed. Natl. Acad. Sci. USA.* **94**, 13193-13197 (1997).

13. Chun, T. W., Engel, D., Berrey, M. M., Shea, T., Corey, L. & Fauci, A. S. Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. *Proceed. Natl. Acad. Sci. USA.* **95**, 8869-8873 (1998).

14. Pantaleo, G., Graziosi, C., Demarest, J. F., Butini, L., Montroni, M., Fox, C. H., Orenstein, J. M., Kotler, D. P. & Fauci, A. S. HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. *Nature* **362**, 355-358 (1993).

15. Embretson, J., Zupancic, M., Ribas, J. L., Burke, A., Racz, P., Tenner-Racz, K. & Haase, A. T. Massive covert infection of helper T-lymphocytes and macrophages by HIV during incubation period of AIDS. *Nature* **362**, 359-362 (1993).

16. Orenstein, J. M., Fox, C. & Wahl, S. M. Macrophages as a source of HIV during opportunistic infections. *Science* **276**, 1857-1861 (1997).

17. Knight, S. C. & Patterson, S. Bone marrow-derived dendritic cells, infection with human immunodeficiency virus, and immunopathology. *Ann. Rev. Immum.* **15**, 593-615 (1997).

18. Chun, T. W., Carruth, L., Finzi, D., Shen, X. F., Digiuseppe, J. A., Taylor, H., Hermankova, M., Chadwick, K., Margolick, J., Quinn, T. C., Kuo, Y. H., Brookmeyer, R., Zeiger, M. A., Barditchcrovo, P. & Siliciano, R. F. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387**, 183-188 (1997).

19. Finzi, D. & Siliciano, R. F. Viral dynamics in HIV-1 infection. *Cell* **93**, 665-671 (1998).

20. Nowak, M. A. & McMichael, A. J. How HIV defeats the immune system. *Sci. Am.* **273**, 58-65 (1995).

21. Wain-Hobson, S. The fastest genome evolution ever described: HIV variation *in situ*. *Curr. Opin. Gen. Develop.* **3**, 878-883 (1993).

22. Sala, M. & Vartanian, J. P. HIV-1 reverse transcriptase - An out-of-the-ordinary enzyme. *Bulletin de L'Institut Pasteur* **96**, 49-63 (1998).

23. Coffin, J. M. HIV population dynamics *in vivo* - implications for genetic variation, pathogenesis, and therapy. *Science* **267**, 483-489 (1995).

24. De Clercq, E. Toward improved anti-HIV chemotherapy: Therapeutic strategies for intervention with HIV infections. *J. Med. Chem.* **38**, 2491-2517 (1995).

25. Stern, A. M. "The development of human immunodeficiency virus type 1 reverse transcriptase inhibitors." in *Chemical and Structural Approaches to Rational Drug Design* (eds. Weiner, D. B. & Williams, W. B.) 35-61 (CRC Press Inc, Boca Raton, Florida, 1995).

26. De Clercq, E. The role of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection. *Antiviral Res.* **38**, 153-179 (1998).

27. De Clercq, E. Perspectives of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection. *Il Farmaco* **54**, 26-45 (1999).

28. Pedersen, O. S. & Pedersen, E. B. Non-nucleoside reverse transcriptase inhibitors: the NNRTI boom. *Antiviral Chem. Chemother.* **10**, 285-314 (1999).

29. Esnouf, R., Ren, J. S., Ross, C., Jones, Y., Stammers, D. & Stuart, D. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nat. Struct. Biol.***2**, 303-308 (1995).

30. Ding, J. P., Das, K., Moereels, H., Koymans, L., Andries, K., Janssen, P. A. J., Hughes, S. H. & Arnold, E. Structure of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. *Nat. Struct. Biol.***2**, 407-415 (1995).

31. Smith, M. B. K., Rouzer, C. A., Taneyhill, L. A., Smith, N. A., Hughes, S. H., Boyer, P. L., Janssen, P. A. J., Moereels, H., Koymans, L., Arnold, E., Ding, J. P., Das, K., Zhang, W. Y., Michejda, C. J. & Smith, R. H. Molecular modeling studies of HIV-1 reverse transcriptase nonnucleoside inhibitors - Total energy of complexation as a predictor of drug placement and activity. *Prot. Sci.***4**, 2203-2222 (1995).

32. Witvrouw, M., Pannecouque, C., Van Laethem, K., Desmyter, J., De Clercq, E. & Vandamme, A. M. Activity of non-nucleoside reverse transcriptase inhibitors against HIV-2 and SIV. *AIDS* **13**, 1477-1483 (1999).

33. Tantillo, C., Ding, J. P., Jacobomolina, A., Nanni, R. G., Boyer, P. L., Hughes, S. H., Pauwels, R., Andries, K., Janssen, P. A. J. & Arnold, E. Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase - Implications for mechanisms of drug inhibition and resistance. *J. Mol. Biol.* **243**, 369-387 (1994).

34. Pedersen, O. S. & Pedersen, E. B. The flourishing syntheses of non-nucleoside reverse transcriptase inhibitors. *Synthesis* **4**, 479-495 (2000).

35. Ng, T. B., Huang, B., Fong, W. P. & Yeung, H. W. Anti-human immunodeficiency virus (anti-HIV) natural products with special emphasis on HIV reverse transcriptase inhibitors. *Life Sci.* **61**, 933-949 (1997).

36. Matthee, G., Wright, A. D. & Konig, G. M. HIV reverse transcriptase inhibitors of natural origin. *Planta Medica.* **65**, 493-506 (1999).

37. Lee, K. H. & Morris-Natschke, S. L. Recent advances in the discovery and development of plant-derived natural products and their analogs as anti-HIV agents. *Pure Appl. Chem.* **71**, 1045-1051 (1999).

38. Jung, M., Lee, S. & Kim, H. Recent studies on natural products as anti-HIV agents. *Curr. Med. Chem.* **7**, 649-661 (2000).

39. Manfredi, K. P., Blunt, J. W., Cardellina, J. H., McMahon, J. B., Pannell, L. L., Cragg, G. M. & Boyd, M. R. Novel alkaloids from the tropical plant *Ancistrocladus abbreviatus* inhibit cell killing in HIV-1 and HIV-2. *J. Med. Chem.* **34**, 3402-3405 (1991).

40. Tan, G. T., Miller, J. F., Kinghorn, A. D., Hughes, S., H, & Pezzuto, J. M. HIV-1 and HIV-2 reverse transcriptases: A comparative study of sensitivity to inhibition by selected natural products. *Biochem. Biophys. Res. Comm.* **185**, 370-378 (1992).

41. Lin, T. S., Schinazi, R., Griffith, B. P., August, E. M., Eriksson, B. F. H., Zheng, D. K., Huang, L. & Prusoff, W. H. Selective inhibition of human immunodeficiency virus type I replication by the (-) but not the (+) enantiomer of Gossypol. *Antimicrob. Agents Chemother.* **33**, 2149-2151 (1989).

42. Keller, P. A., Leach, S. P., Birch, C., Tyssen, D. & Griffith, R. *J. Med. Chem. (submitted)* (2000).

43. Royer, R. E., Deck, L. M., Vanderjagt, T. J., Martinez, F. J., Mills, R. G., Young, S. A. & Vanderjagt, D. L. Synthesis and anti-HIV activity of 1,1'-dideoxygossypol and related compounds. *J. Med. Chem.* **38**, 2427-2432 (1995).

44. Xie, L., Xie, J. X., Kashiwada, Y., Cosentino, L. M., Liu, S. H., Pai, R. B., Cheng, Y. C. & Lee, K. H. Anti-AIDS (Acquired Immune Deficiency Syndrome) agents .17. New brominated hexahydroxybiphenyl derivatives as potent anti-HIV agents. *J. Med. Chem.* **38**, 3003-3008 (1995).

45. Gustafson, K. R., Cardellina, J. H., McMahon, J. B., Pannell, L. K., Cragg, G. M. & Boyd, M. R. The Peltatols, novel HIV-inhibitory catechol derivatives from *Pothomorphe peltata. J. Org. Chem.* **57**, 2809-2811 (1992).

46. Decosterd, L. A., Parsons, I. C., Gustafson, K. R., Cardellina, J. H., McMahon, J. B., Cragg, G. M., Murata, Y., Pannell, L. K., Steiner, J. R., Clardy, J. & Boyd, M. R. HIV inhibitory natural products .11. Structure, absolute stereochemistry, and synthesis of Conocurvone, a potent, novel HIV-inhibitory naphthoquinone trimer from a *Conospermum Sp. J. Am. Chem. Soc.* **115**, 6673-6679 (1993).

47. Boyd, M. R., Hallock, Y. F., Manfredi, K. P., Blunt, J. W., McMahon, J. B., Buckheit, R. W., Bringmann, G., Schaffer, M., Cragg, G. M., Thomas, D. W., Jato, J. G. & Cardellina, J. H. Anti-HIV michellamines from *Ancistrocladus korupensis. J. Med. Chem.* **37**, 1740-1745 (1994).

48. Supko, J. G. & Malspeis, L. Pharmacokinetics of Michellamine B, a naphthylisoquinoline alkaloid with in vitro activity against human immunodeficiency virus types 1 and 2, in the mouse and dog. *Antimicrob. Agents Chemother.* **39**, 9-14 (1995).

49. McMahon, J. B., Currens, M. J., Gulakowski, R. J., Buckheit, R. W., Lackmansmith, C., Hallock, Y. F. & Boyd, M. R. Michellamine B, a novel plant alkaloid, inhibits human immunodeficiency virus-induced cell killing by at least two distinct mechanisms. *Antimicrob. Agents Chemother.* **39**, 484-488 (1995).

50. White, E. L., Chao, W. R., Ross, L. J., Borhani, D. W., Hobbs, P. D., Upender, V. & Dawson, M. I. Michellamine alkaloids inhibit protein kinase C. *Arch. Biochem. Biophys.* **365**, 25-30 (1999).

51. White, E. L., Ross, L. J., Hobbs, P. D., Upender, V. & Dawson, M. I. Antioxidant activity of michellamine alkaloids. *Anticancer Res.*. **19**, 1033-1035 (1999).

52. Bringmann, G., Zagst, R., Reuscher, H. & Assi, L. A. Ancistrobrevine B, the first naphthylisoquinoline alkaloid with a 5,8'-coupling site, and related compounds from *Ancistrocladus abbreviatus. Phytochemistry* **31**, 4011-4014 (1992).

53. Bringmann, G., Gotz, R., Keller, P. A., Walter, R., Henschel, P., Schaffer, M., Stablein, M., Kelly, T. R. & Boyd, M. R. Acetogenic isoquinoline alkaloids .66. First total synthesis of Korupensamines A and B. *Heterocycles* **39**, 503-508 (1994).

54. Bringmann, G., Zagst, R., Schaffer, M., Hallock, Y. F., Cardellina, J. H. & Boyd, M. R. The absolute configuration of Michellamine-B, a 'dimeric', anti-HIV-active naphthylisoquinoline alkaloid. *Angew. Chem. Int. Ed. Engl.* **32**, 1190-1191 (1993).

55. Kalsi, P. S. *Sterochemistry: conformation and mechanism* (Wiley Eastern Limited, New Delhi, India, 1990).

56. "Sterochemistry, hydrocarbons, halo compounds, oxygen compounds." in *Comprehensive Organic Chemistry: The Synthesis and Reactions of Organic Compounds* (ed. Stoddart, J. F.) (Pergamon Press, Oxford, England, 1979).

57. "Stereochemistry." in *Nomenclature of Organic Chemistry* (eds. Rigaudy, J. & Klesney, S. P.) (Pergamon Press, Oxford, England, 1979).

58. Hallock, Y. F., Manfredi, K. P., Dai, J. R., Cardellina, J. H., Gulakowski, R. J., McMahon, J. B., Schaffer, M., Stahl, M., Gulden, K. P., Bringmann, G., Francois, G. & Boyd, M. R. HIV-Inhibitory natural products .36. Michellamines D-F, new HIV-inhibitory dimeric naphthylisoquinoline alkaloids, and Korupensamine E, a new antimalarial monomer, from *Ancistrocladus korupensis*. *J. Nat. Prod.* **60**, 677-683 (1997).

59. Bringmann, G., Harmsen, S., Holenz, J., Geuder, T., Gotz, R., Keller, P. A., Walter, R., Hallock, Y. F., Cardellina, J. H. & Boyd, M. R. Acetogenic isoquinoline alkaloids .69. 'Biomimetic' oxidative dimerization of korupensamine A - Completion of the first total synthesis of Michellamines A, B, and C. *Tetrahedron* **50**, 9643-9648 (1994).

60. Hallock, Y. F., Manfredi, K. P., Blunt, J. W., Cardellina, J. H., Schaffer, M., Gulden, K. P., Bringmann, G., Lee, A. Y., Clardy, J., Francois, G. & Boyd, M. R. Korupensamines A-D, novel antimalarial alkaloids from *Ancistrocladus korupensis*. *J. Org. Chem.* **59**, 6349-6355 (1994).

61. Hoye, T. R. & Mi, L. Total syntheses of Korupensamine C and Ancistrobrevine B. *Tet. Lett.* **37**, 3097-3098 (1996).

62. Bringmann, G., Gotz, R., Harmsen, S., Holenz, J. & Walter, R. Acetogenic isoquinoline alkaloids .82. Biomimetic total synthesis of Michellamines A-C. *Liebigs Ann.* 2045-2058 (1996).

63. Hoye, T. R., Chen, M. Z., Mi, L. & Priest, O. P. Total synthesis of Michellamines A-C - Important anti-HIV agents. *Tet. Lett.* **35**, 8747-8750 (1994).

64. Hoye, T. R., Chen, M. Z., Hoang, B., Mi, L. & Priest, O. P. Total synthesis of Michellamines A-C, Korupensamines A-D, and Ancistrobrevine B. *J. Org. Chem.* **64**, 7184-7201 (1999).

65. Hobbs, P. D., Upender, V., Liu, J. W., Pollart, D. J., Thomas, D. W. & Dawson, M. I. The first stereospecific synthesis of Michellamine B. *Chem. Comm.* 923-924 (1996).

66. Hobbs, P. D., Upender, V. & Dawson, M. I. Stereospecific syntheses of Michellamines A and C. *Synlett*, 965 ff. (1997).

67. Kelly, T. R., Garcia, A., Lang, F. R., Walsh, J. J., Bhaskar, K. V., Boyd, M. R., Gotz, R., Keller, P. A., Walter, R. & Bringmann, G. Convergent total synthesis of the michellamines. *Tet. Lett.* **35**, 7621-7624 (1994).

68. Bringmann, G., Gotz, R., Keller, P. A., Walter, R., Boyd, M. R., Lang, F. R., Garcia, A., Walsh, J. J., Tellitu, I., Bhaskar, K. V. & Kelly, T. R. A convergent total synthesis of the michellamines. *J. Org. Chem.* **63**, 1090-1097 (1998).

69. Bringmann, G., Ortmann, T., Feineis, D., Peters, E. M. & Peters, K. A new central binaphthalene building block for michellamine syntheses. *Synthesis* **3**, 383-388 (2000).

70. Upender, V., Pollart, D. J., Liu, J., Hobbs, P. D., Olsen, C., Chao, W. R., Bowden, B., Crase, J. L., Thomas, D. W., Pandey, A., Lawson, J. A. & Dawson, M. I. The synthesis and biological activity of two analogs of the anti-HIV alkaloid Michellamine B. *Journal of Heterocyclic Chemistry* **33**, 1371-1384 (1996).

71. Zhang, H. P., Zembower, D. E. & Chen, Z. D. Structural analogues of the michellamine anti-HIV agents - Importance of the tetrahydroisoquinoline rings for biological activity. *Bioorg. Med. Chem. Lett.* **7**, 2687-2690 (1997).

72. Bringmann, G. Mono- and dimeric naphthylisoquinoline alkaloids - Pharmaceutically and structurally exciting natural heterocycles with axial chirality. *Bulletin des Societes Chimiques Belges* **105**, 601-613 (1996).

73. Bringmann, G., Ortmann, T., Zagst, R., Schöner, B., Assi, L. A. & Burschka, C. (±)-Dioncophyllacine A, a naphthylisoquinoline alkaloid with a 4-methoxy substituent from the leaves of *Triphyophyllum peltatum*. *Phytochemistry* **31**, 4015-4018 (1992).

74. Bringmann, G., Rübenacker, M., Weirich, R. & Assi, L. A. Dioncophylline C from the roots of *Triphyophyllum peltatum*, the first 5,1'-coupled dioncophyllaceae alkaloid. *Phytochemistry* **31**, 4019-4024 (1992).

75. Bringmann, G., Holenz, J., Weirich, R., Rubenacker, M., Funke, C., Boyd, M. R., Gulakowski, R. J. & Francois, G. First synthesis of the antimalarial naphthylisoquinoline alkaloid Dioncophylline C, and its unnatural anti-HIV dimer, Jozimine C. *Tetrahedron* **54**, 497-512 (1998).

76. Bringmann, G., Saeb, W., Koppler, D. & Francois, G. Jozimine A ('dimeric' Dioncophylline A), a non-natural michellamine analog with high antimalarial activity. *Tetrahedron* **52**, 13409-13418 (1996).

77. Greer, J., Erickson, J. W., Baldwin, J. J. & Varney, M. D. Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* **37**, 1035-1054 (1994).

78. Boyd, D. B. "Drug design." in *Encyclopedia of Computational Chemistry* (eds. Schleyer, P. v. R. *et al.*) 795-804 (John Wiley & Sons Ltd, Chichester, UK, 1998).

79. Sarafianos, S. G., Das, K., Ding, J. P., Hsiou, Y., Hughes, S. H. & Arnold, E. "Current perspectives on mechanisms of HIV-1 reverse transcriptase inhibition by nonucleoside inhibitors." in *Anti-infectives: Recent Advances in Chemistry and Structure-Activity Relationships* (eds. Bentley, P. H. & O'Hanlon, P. J.) 328-334 (The Royal Society of Chemistry, Cambridge, 1997).

80. Stanforth, S. P. Catalytic cross-coupling reactions in biaryl synthesis. *Tetrahedron* **54**, 263-303 (1998).

81. Fanta, P. E. The Ullmann synthesis of biaryls. *Synthesis*, 9-21 (1974).

82. Negishi, E., King, A. O. & Okukado, N. Selective carbon-carbon bond formation via transition metal catalysis. 3. A highly selective synthesis of unsymmetrical biaryls and diarylmethanes by the nickel- or palladium-catalyzed reaction of aryl- and benzylzinc derivatives with aryl halides. *J. Org. Chem.* **42**, 1821 (1977).

83. Hoye, T. R. & Chen, M. Z. Studies of palladium-catalyzed cross-coupling reactions for preparation of highly hindered biaryls relevant to the Korupensamine/Michellamine problem. *J. Org. Chem.* **61**, 7940-7942 (1996).

84. Stille, J. K. The palladium-catalyzed cross-coupling reactions of organotin reagents with organic electrophiles. *Angew. Chem. Int. Ed. Engl.* **25**, 508-524 (1986).

85. Suzuki, A. New synthetic transformations via organoboron compounds. *Pure Appl. Chem.* **66**, 213-222 (1994).

86. Miyaura, N. & Suzuki, A. Palladium-catalyzed cross-coupling reactions of organoboron compounds. *Chem. Rev.* **95**, 2457-2483 (1995).

87. Suzuki, A. Recent advances in the cross-coupling reactions of organoboron derivatives with organic electrophiles, 1995-1998. *J. Organomet. Chem.* **576**, 147-168 (1999).

88. Mitchell, T. N. Palladium-catalysed reactions of organotin compounds. *Synthesis*, 805-815 (1992).

89. Duncton, M. A. J. & Pattenden, G. The intramolecular Stille reaction. *J. Chem. Soc. Perkin Trans. I.* 1235-1246 (1999).

90. Farina, V., Kapadia, S., Krishnan, B., Wang, C. J. & Liebeskind, L. S. On the nature of the 'copper effect' in the Stille cross-coupling. *J. Org. Chem.* **59**, 5905-5911 (1994).

91. Roth, G. P., Farina, V., Liebeskind, L. S. & Penacabrera, E. Optimized stille coupling reactions catalyzed by palladium on carbon with CuI as cocatalyst. *Tet. Lett.* **36**, 2191-2194 (1995).

92. Tamayo, N., M., E. A., C., P. M., Farina, F. & Noheda, P. Synthesis of anthraquinone derivatives by palladium-catalysed coupling of triflates with stannanes. *Tet. Lett.* **31**, 5189-5192 (1990).

93. Kelly, T. R. & Kim, M. H. Synthesis of schumanniophytine and isoschumanniophytine. *J. Org. Chem.* **57**, 1593-1597 (1992).

94. Ohe, T., Miyaura, N. & Suzuki, A. Palladium-catalyzed cross-coupling reaction of organoboron compounds with organic triflates. *J. Org. Chem.* **58**, 2201-2208 (1993).

95. Watanabe, T., Miyaura, N. & Suzuki, A. Synthesis of sterically hindered biaryls via the palladium-catalyzed cross-coupling reaction of arylboronic acids or their esters with haloarenes. *Synlett* **8**, 207-210 (1992).

96. Zhang, H. C., Kwong, F. Y., Tian, Y. & Chan, K. S. Base and cation effects on the suzuki cross-coupling of bulky arylboronic acid with halopyridines - Synthesis of pyridylphenols. *J. Org. Chem.* **63**, 6886-6890 (1998).

97. Moreno-Manas, M., Perez, M. & Pleixats, R. Palladium-catalyzed suzuki-type self-coupling of arylboronic acids - A mechanistic study. *J. Org. Chem.* **61**, 2346-2351 (1996).

98. Miller, R. B. & Dugar, S. Stoichiometric synthesis of unsymmetrical mononitrobiphenyls via the palladium-catalyzed cross-coupling of arylboronic acids with aryl bromides. *Organometallics* **3**, 1261-1263 (1984).

99. Andersen, N. G., Maddaford, S. P. & Keay, B. A. A modified in situ Suzuki cross-coupling of haloarenes for the preparation of C_2-symmetrical biaryls. *J. Org. Chem.* **61**, 9556-9559 (1996).

100. Benbow, J. W. & Martinez, B. L. Biaryl formation using the Suzuki protocol - Considerations of base, halide, and protecting group. *Tet. Lett.* **37**, 8829-8832 (1996).

101. Giroux, A., Han, Y. X. & Prasit, P. One pot biaryl synthesis via *in situ* boronate formation. *Tet. Lett.* **38**, 3841-3844 (1997).

102. Ishiyama, T., Itoh, Y., Kitano, T. & Miyaura, N. Synthesis of arylboronates via the palladium(0)-catalyzed cross-coupling reaction of tetra(alkoxo)diborons with aryl triflates. *Tet. Lett.* **38**, 3447-3450 (1997).

103. Ishiyama, T., Murata, M. & Miyaura, N. Palladium(0)-catalyzed cross-coupling reaction of alkoxydiboron with haloarenes - A direct procedure for arylboronic esters. *J. Org. Chem.* **60**, 7508-7510 (1995).

104. Kong, K. C. & Cheng, C. H. Facile aryl-aryl exchange between the palladium center and phosphine ligands in palladium(II) complexes. *J. Am. Chem. Soc.* **113**, 6313-6315 (1991).

105. O'Keefe, D. F., Dannock, M. C. & Marcuccio, S. M. Palladium catalysed coupling of halobenzenes with arylboronic acids: Role of the triphenylphosphine ligand. *Tet. Lett.* **33**, 6679-6680 (1992).

106. Smith, K. A., Campi, E. M., Jackson, W. R., Marcuccio, S., Naeslund, C. G. M. & Deacon, G. B. High yields of symmetrical biaryls from palladium catalysed homocoupling of arylboronic acids under mild conditions. *Synlett*, 131 ff. (1997).

107. Koyama, H. & Kamikawa, T. Total syntheses of $O^{4,9}$-dimethyl stealthins A and C. *Tet. Lett.* **38**, 3973-3976 (1997).

108. Marck, G., Villiger, A. & Buchecker, R. Aryl couplings with heterogeneous palladium catalysts. *Tet. Lett.* **35**, 3277-3280 (1994).

109. Owton, W. M., Gallagher, P. T. & Juanmontesinos, A. Tert-butyl 3-carboxyethyl-3-phosphonodiethylpropionate - A novel reagent for Stobbe-like condensations. *Synth. Comm.* **23**, 2119-2125 (1993).

110. Aliprantis, A. O. & Canary, J. W. Observation of catalytic intermediates in the suzuki reaction by electrospray mass spectrometry. *J. Am. Chem. Soc.* **116**, 6985-6986 (1994).

111. Hayashi, T., Konishi, M., Kobori, Y., Kumada, M., Higuchi, T. & Hirotsu, K. Dichloro[1.1'-bis(diphenylphosphino)ferrocene]palladium(II): An effective catalyst for cross-coupling of secondary and primary alkyl Grignard and alkylzinc reagents with organic halides. *J. Am. Chem. Soc.* **106**, 158-163 (1984).

112. Coulson, D. R. Tetrakis(triphenylphosphine) palladium (0). *Inorg. Synth.***13**, 121-123 (1972).

113. Malan, C. & Morin, C. A concise preparation of 4-borono-L-phenylalanine (L-BPA) from L-phenylalanine. *J. Org. Chem.* **63**, 8019-8020 (1998).

114. Segelstein, B. E., Butler, T. W. & Chenard, B. L. Equilibration of the oxidative addition product of tetrakis(triphenylphosphine) palladium and electron-rich aryl halides leads to product scrambling in the stille reaction. *J. Org. Chem.* **60**, 12-13 (1995).

115. Hitchcock, S. A., Mayhugh, D. R. & Gregory, G. S. Selectivity in palladium(0)-catalyzed cross-coupling reactions - Application to a tandem Stille reaction. *Tet. Lett.* **36**, 9085-9088 (1995).

116. Brimble, M. A., Neville, D. & Duncalf, L. J. Synthesis of a dimeric pyranonaphthoquinone via a novel double furofuran annulation strategy. *Tet. Lett.* **39**, 5647-5650 (1998).

117. Lan, A. J. Y., Heuckeroth, R. O. & Mariano, P. S. Electron-transfer-induced photocyclization reactions of arene-iminium salt systems. Effects of cation diradical deprotonation and desilylation on the nature and efficiencies of reaction pathways followed. *J. Am. Chem. Soc.* **109**, 2738-2745 (1987).

118. AMRAD Operations (AMRAD, 2000). http://www.amrad.com.au.

119. Paitayatat, S., Tarnchompoo, B., Thebtaranonth, Y. & Yuthavong, Y. Correlation of antimalarial activity of Artemisinin derivatives with binding affinity with ferroprotoporphyrin IX. *J. Med. Chem.* **40**, 633-638 (1997).

120. Baltimore, D. Viral RNA-dependent DNA polymerase. *Nature* **226**, 1209-1211 (1970).

121. Temin, H. M. & Mizutani, S. RNA-dependant DNA polymerase in virions of Rous Sarcoma Virus. *Nature* **226**, 1211-1213 (1970).

122. Telesnitsky, A. & Goff, S. P. "Reverse transcriptase and the generation of retroviral DNA." in *Retroviruses* (eds. Coffin, J. M., Hughes, S. H. & Varmus, H. E.) 121-160 (Cold Spring Harbor Laboratory Press, New York, 1997).

123. Furman, P. A., Painter, G. R. & Anderson, K. S. An analysis of the catalytic cycle of HIV-1 reverse transcriptase: Opportunities for chemotherapeutic intervention based on enzyme inhibition. *Curr. Pharm. Des.* **6**, 547-567 (2000).

124. Nanni, R. G., Ding, J. P., Jacobo-Molina, A., Hughes, S. H. & Arnold, E. Review of HIV-1 reverse transcriptase three-dimensional structure: Implications for drug design. *Perspect. Drug Discov. Design* **1**, 129-150 (1993).

125. Arts, E. J. & Wainberg, M. A. Human immunodeficiency virus type 1 reverse transcriptase and early events in reverse transcription. *Adv. Virus Res.* **46**, 97-163 (1996).

126. Litvak, S. *Retroviral Reverse Transcriptases* (R. G. Landes Company, Heidelberg, Germany, 1996).

127. Gotte, M., Li, X. G. & Wainberg, M. A. HIV-1 reverse transcription: A brief overview focused on structure-function relationships among molecules involved in initiation of the reaction. *Arch. Biochem. Biophys.*. **365**, 199-210 (1999).

128. Jonckheere, H., Anne, J. & De Clercq, E. The HIV-1 reverse transcription (RT) process as target for RT inhibitors. *Med. Res. Rev.* **20**, 129-154 (2000).

129. Hostomsky, Z., Hughes, S. H., Goff, S. P. & Legrice, S. F. J. Redesignation of the RNase D activity associated with retroviral reverse transcriptase as RNase H*. *J. Virol.* **68**, 1970-1971 (1994).

130. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783-1790 (1992).

131. Harris, D., Lee, R., Misra, H. S., Pandey, P. K. & Pandey, V. N. The p51 subunit of human immunodeficiency virus type 1 reverse transcriptase is essential in loading the p66 subunit on the template primer. *Biochemistry* **37**, 5903-5908 (1998).

132. Jacobo Molina, A., Ding, J. P., Nanni, R. G., Clark, A. D., Lu, X. D., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., Hizi, A., Hughes, S. H. & Arnold, E. Crystal structure of human immunodeficiency virus type-1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proceed. Natl. Acad. Sci. USA.* **90**, 6320-6324 (1993).

133. Ren, J. S., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., Stuart, D. & Stammers, D. High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat. Struct. Biol.***2**, 293-302 (1995).

134. Ding, J., Das, K., Tantillo, C., Zhang, W., Clark, A. D., Jessen, S., Lu, X., Hsiou, Y., Jacobomolina, A., Andries, K., Pauwels, R., Moereels, H., Koymans, L., Janssen, P. A. J., Smith, R. H., Koepke, M. K., Michejda, C. J., Hughes, S. H. & Arnold, E. Structure of HIV-1 reverse transcriptase in a complex with the non-nucleoside inhibitor α-APA R 95845 at 2.8 Å resolution. *Structure* **3**, 365-379 (1995).

135. Rodgers, D. W., Gamblin, S. J., Harris, B. A., Ray, S., Culp, J. S., Hellmig, B., Woolf, D. J., Debouck, C. & Harrison, S. C. The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. *Proceed. Natl. Acad. Sci. USA.* **92**, 1222-1226 (1995).

136. Huang, H. F., Chopra, R., Verdine, G. L. & Harrison, S. C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase - Implications for drug resistance. *Science* **282**, 1669-1675 (1998).

137. Bahar, I., Erman, B., Jernigan, R. L., Atilgan, A. R. & Covell, D. G. Collective motions in HIV-1 reverse transcriptase: Examination of flexibility and enzyme function. *J. Mol. Biol.* **285**, 1023-1037 (1999).

138. Ren, J. S., Esnouf, R., Hopkins, A., Ross, C., Jones, Y., Stammers, D. & Stuart, D. The structure of HIV-1 reverse transcriptase complexed with 9-chloro-TIBO: lessons for inhibitor design. *Structure* **3**, 915-926 (1995).

139. Ding, J. P., Das, K., Hsiou, Y., Sarafianos, S. G., Clark, A. D., Jacobo-Molina, A., Tantillo, C., Hughes, S. H. & Arnold, E. Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *J. Mol. Biol.* **284**, 1095-1111 (1998).

140. May, A. C. W. Pairwise iterative superposition of distantly related proteins and assessment of the significance of 3-D structural similarity. *Prot. Eng.* **9**, 1093-1101 (1996).

141. May, A. C. W. & Johnson, M. S. Improved genetic algorithm-based protein structure comparisons - Pairwise and multiple superpositions. *Prot. Eng.* **8**, 873-882 (1995).

142. Ren, J. S., Esnouf, R. M., Hopkins, A. L., Warren, J., Balzarini, J., Stuart, D. I. & Stammers, D. K. Crystal structures of HIV-1 reverse transcriptase in complex with carboxanilide derivatives. *Biochemistry* **37**, 14394-14403 (1998).

143. Jäger, J., Smerdon, S. J., Wang, J. M., Boisvert, D. C. & Steitz, T. A. Comparison of three different crystal forms shows HIV-1 reverse transcriptase displays an internal swivel motion. *Structure* **2**, 869-876 (1994).

144. Das, K., Ding, J. P., Hsiou, Y., Clark, A. D., Moereels, H., Koymans, L., Andries, K., Pauwels, R., Janssen, P. A. J., Boyer, P. L., Clark, P., Smith, R. H., Smith, M. B. K., Michejda, C. J., Hughes, S. H. & Arnold, E. Crystal structures of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. *J. Mol. Biol.* **264**, 1085-1100 (1996).

145. Lesk, A. M. Extraction of well-fitting substructures - Root-mean-square deviation and the difference distance matrix. *Folding Des.* **2**, S12-S14 (1997).

146. Lesk, A. M. Extraction of geometrically similar substructures - Least-squares and Chebyshev fitting and the difference distance matrix. *Proteins* **33**, 320-328 (1998).

147. Keller, P. A., Leach, S. P., Luu, T. T. T., Titmuss, S. J. & Griffith, R. Development of computational and graphical tools for the analysis of movement and flexibility in large molecules. *J. Mol. Graph. Mod.* **18**, 235-241 (2000).

148. Paakkonen, K., Annila, A., Sorsa, T., Pollesello, P., Tilgmann, C., Kilpelainen, I., Karisola, P., Ulmanen, I. & Drakenberg, T. Solution structure and main chain dynamics of the regulatory domain (residues 1-91) of human cardiac troponin C. *J. Biol. Chem.* **273**, 15633-15638 (1998).

149. Nichols, W. L., Rose, G. D., Teneyck, L. F. & Zimm, B. H. Rigid domains in proteins - An algorithmic approach to their identification. *Proteins* **23**, 38-48 (1995).

150. Akke, M., Forsen, S. & Chazin, W. J. Solution structure of $(Cd^{2+})_1$-calbindin D_{9k} reveals details of the stepwise structural changes along the apo→$(Ca^{2+})_1^{II}$→$(Ca^{2+})_2^{I,II}$ binding pathway. *J. Mol. Biol.* **252**, 102-121 (1995).

151. Leach, S. (1999). "ProFlex: A computational tool for the investigation and analysis of movement and flexibility in large molecules," Honours, University of Wollongong, Wollongong.

152. InsightII® 97.0 (Molecular Simulations Inc., San Diego, CA, USA, 1997).

153. Excel® 97 (Microsoft Corporation, 1997).

154. Patel, P. H., Jacobomolina, A., Ding, J. P., Tantillo, C., Clark, A. D., Raag, R., Nanni, R. G., Hughes, S. H. & Arnold, E. Insights into DNA polymerization mechanisms from structure and function analysis of HIV-1 reverse transcriptase. *Biochemistry* **34**, 5351-5363 (1995).

155. Divita, G., Muller, B., Immendorfer, U., Gautel, M., Rittinger, K., Restle, T. & Goody, R. S. Kinetics of interaction of HIV reverse transcriptase with primer/template. *Biochemistry* **32**, 7966-7971 (1993).

156. Kati, W. M., Johnson, K. A., Jerva, L. F. & Anderson, K. S. Mechanism and fidelity of HIV reverse transcriptase. *J. Biol. Chem.* **267**, 25988-25997 (1992).

157. Hsieh, J. C., Zinnen, S. & Modrich, P. Kinetic mechanism of the DNA-dependent DNA polymerase activity of human immunodeficiency virus reverse transcriptase. *J. Biol. Chem.* **268**, 24607-24613 (1993).

158. Reardon, J. E. Human immunodeficiency virus reverse transcriptase: Steady-state and pre-steady-state kinetics of nucleotide incorporation. *Biochemistry* **31**, 4473-4479 (1992).

159. Huang, H. F., Harrison, S. C. & Verdine, G. L. Trapping of a catalytic HIV reverse transcriptase center•template : primer complex through a disulfide bond. *Chem. Biol.* **7**, 355-364 (2000).

160. Sarafianos, S. G., Das, K., Ding, J. P., Boyer, P. L., Hughes, S. H. & Arnold, E. Touching the heart of HIV-1 drug resistance: the fingers close down on the dNTP at the polymerase active site. *Chem. Biol.*. **6**, R137-R146 (1999).

161. Kaushik, N., Rege, N., Yadav, P. N. S., Sarafianos, S. G., Modak, M. J. & Pandey, V. N. Biochemical analysis of catalytically crucial aspartate mutants of human immunodeficiency virus type 1 reverse transcriptase. *Biochemistry* **35**, 11536-11546 (1996).

162. Jonckheere, H., De Clercq, E. & Anne, J. Fidelity analysis of HIV-1 reverse transcriptase mutants with an altered amino-acid sequence at residues Leu74, Glu89, Tyr115, Tyr183 and Met184. *Eur. J. Biochem.* **267**, 2658-2665 (2000).

163. Harris, D., Yadav, P. N. S. & Pandey, V. N. Loss of polymerase activity due to Tyr to Phe substitution in the YMDD motif of human immunodeficiency virus type-1 reverse transcriptase is compensated by Met to Val substitution within the same motif. *Biochemistry* **37**, 9630-9640 (1998).

164. Pandey, V. N., Kaushik, N., Rege, N., Sarafianos, S. G., Yadav, P. N. S. & Modak, M. J. Role of methionine 184 of human immunodeficiency virus type-1 reverse transcriptase in the polymerase function and fidelity of DNA synthesis. *Biochemistry* **35**, 2168-2179 (1996).

165. Harris, D., Kaushik, N., Pandey, P. K., Yadav, P. N. S. & Pandey, V. N. Functional analysis of amino acid residues constituting the dNTP binding pocket of HIV-1 reverse transcriptase. *J. Biol. Chem.* **273**, 33624-33634 (1998).

166. Boyer, P. L., Ferris, A. L. & Hughes, S. H. Cassette mutagenesis of the reverse transcriptase of human immunodeficiency virus type 1. *J. Virol.* **66**, 1031-1039 (1992).

167. Spence, R. A., Kati, W. M., Anderson, K. S. & Johnson, K. A. Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. *Science* **267**, 988-993 (1995).

168. Arnold, E., Ding, J. P., Hughes, S. H. & Hostomsky, Z. Structures of DNA and RNA polymerases and their interactions with nucleic acid substrates. *Curr. Opin. Struct. Biol.* **5**, 27-38 (1995).

169. Jacques, P. S., Wohrl, B. M., Ottmann, M., Darlix, J. L. & Legrice, S. F. J. Mutating the primer grip of p66 HIV-1 reverse transcriptase implicates tryptophan-229 in template-primer utilization. *J. Biol. Chem.* **269**, 26472-26478 (1994).

170. Ghosh, M., Jacques, P. S., Rodgers, D. W., Ottman, M., Darlix, J. L. & Legrice, S. F. J. Alterations to the primer grip of p66 HIV-1 reverse transcriptase and their consequences for template-primer utilization. *Biochemistry* **35**, 8553-8562 (1996).

171. Powell, M. D., Ghosh, M., Jacques, P. S., Howard, K. J., Legrice, S. F. J. & Levin, J. G. Alanine-scanning mutations in the primer grip of p66 HIV-1 reverse transcriptase result in selective loss of RNA priming activity. *J. Biol. Chem.* **272**, 13262-13269 (1997).

172. Palaniappan, C., Wisniewski, M., Jacques, P. S., Legrice, S. F. J., Fay, P. J. & Bambara, R. A. Mutations within the primer grip region of HIV-1 reverse transcriptase result in loss of RNase H function. *J. Biol. Chem.* **272**, 11157-11164 (1997).

173. Ghosh, M., Williams, J., Powell, M. D., Levin, J. G. & Legrice, S. F. J. Mutating a conserved motif of the HIV-1 reverse transcriptase palm subdomain alters primer utilization. *Biochemistry* **36**, 5758-5768 (1997).

174. Fedoroff, O. Y., Ge, Y. & Reid, B. R. Solution structure of R(GAGGACUG)-D(CAGTCCTC) hybrid - Implications for the initiation of HIV-1 (+)-strand synthesis. *J. Mol. Biol.* **269**, 225-239 (1997).

175. Szyperski, T., Gotte, M., Billeter, M., Perola, E., Cellai, L., Heumann, H. & Wuthrich, K. NMR structure of the chimeric hybrid duplex r(gcaguggc).r(gcca)d(CTGC) comprising the tRNA-DNA junction formed during initiation of HIV-1 reverse transcription. *J. Biomol. NMR.* **13**, 343-355 (1999).

176. Elgavish, T., VanLoock, M. S. & Harvey, S. C. Exploring three-dimensional structures of the HIV-1 RNA/tRNA^{Lys3} initiation complex. *J. Mol. Biol.* **285**, 449-453 (1999).

177. Wisniewski, M., Palaniappan, C., Fu, Z. P., Le Grice, S. F. J., Fay, P. & Bambara, R. A. Mutations in the primer grip region of HIV reverse transcriptase can increase replication fidelity. *J. Biol. Chem.* **274**, 28175-28184 (1999).

178. Lanchy, J. M., Keith, G., Legrice, S. F. J., Ehresmann, B., Ehresmann, C. & Marquet, R. Contacts between reverse transcriptase and the primer strand govern the transition from initiation to elongation of HIV-1 reverse transcription. *J. Biol. Chem.* **273**, 24425-24432 (1998).

179. Pelemans, H., Esnouf, R., De Clercq, E. & Balzarini, J. Mutational analysis of Trp-229 of human immunodeficiency virus type 1 reverse transcriptase (RT) identifies this amino acid residue as a prime target for the rational design of new non-nucleoside RT inhibitors. *Mol. Pharm.* **57**, 954-960 (2000).

180. Boyer, P. L., Ferris, A. L. & Hughes, S. H. Mutational analysis of the fingers domain of human immunodeficiency virus type 1 reverse transcriptase. *J. Virol.* **66**, 7533-7537 (1992).

181. Boyer, P. L., Ferris, A. L., Clark, P., Whitmer, J., Frank, P., Tantillo, C., Arnold, E. & Hughes, S. H. Mutational analysis of the fingers and palm subdomains of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase. *J. Mol. Biol.* **243**, 472-483 (1994).

182. Chao, S. F., Chan, V. L., Juranka, P., Kaplan, A. H., Swanstrom, R. & Hutchison, C. A. Mutational sensitivity patterns define critical residues in the palm subdomain of the reverse transcriptase of human immunodeficiency virus type 1. *Nucleic Acids Res.* **23**, 803-810 (1995).

183. Kew, Y., Olsen, L. R., Japour, A. J. & Prasad, V. R. Insertions into the β3-β4 hairpin loop of HIV-1 reverse transcriptase reveal a role for fingers subdomain in processive polymerization. *J. Biol. Chem.* **273**, 7529-7537 (1998).

184. Kim, B., Ayran, J. C., Sagar, S. G., Adman, E. T., Fuller, S. M., Tran, N. H. & Horrigan, J. New human immunodeficiency virus, type 1 reverse transcriptase (HIV-1 RT) mutants with increased fidelity of DNA synthesis - Accuracy, template binding, and processivity. *J. Biol. Chem.* **274**, 27666-27673 (1999).

185. Kim, B., Hathaway, T. R. & Loeb, L. A. Fidelity of mutant HIV-1 reverse transcriptases - Interaction with the single-stranded template influences the accuracy of DNA synthesis. *Biochemistry* **37**, 5831-5839 (1998).

186. Sarafianos, S. G., Pandey, V. N., Kaushik, N. & Modak, M. J. Site-directed mutagenesis of arginine 72 of HIV-1 reverse transcriptase - Catalytic role and inhibitor sensitivity. *J. Biol. Chem.* **270**, 19729-19735 (1995).

187. Kaushik, N., Harris, D., Rege, N., Modak, M. J., Yadav, P. N. S. & Pandey, V. N. Role of glutamine-151 of human immunodeficiency virus type-1 reverse transcriptase in RNA-directed DNA synthesis. *Biochemistry* **36**, 14430-14438 (1997).

188. Kaushik, N., Talele, T. T., Pandey, P. K., Harris, D., Yadav, P. N. S. & Pandey, V. N. Role of glutamine 151 of human immunodeficiency virus type-1 reverse transcriptase in substrate selection as assessed by site-directed mutagenesis. *Biochemistry* **39**, 2912-2920 (2000).

189. Sarafianos, S. G., Pandey, V. N., Kaushik, N. & Modak, M. J. Glutamine 151 participates in the substrate dNTP binding function of HIV-1 reverse transcriptase. *Biochemistry* **34**, 7207-7216 (1995).

190. Davies, J. F., Hostomsky, Z., Jordan, S. R. & Matthews, D. A. Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science* **252**, 88-95 (1991).

191. Hughes, S. H., Arnold, E. & Hostomsky, Z. "RNase H of retroviral reverse transcriptases." in *Ribonucleases H* (eds. Crouch, R. J. & Toulme, J. J.) 195-224 (Paris, 1998).

192. Gotte, M., Maier, G., Gross, H. J. & Heumann, H. Localization of the active site of HIV-1 reverse transcriptase-associated RNase H domain on a DNA template using site-specific generated hydroxyl radicals. *J. Biol. Chem.* **273**, 10139-10146 (1998).

193. Destefano, J. J., Wu, W. M., Seehra, J., McCoy, J., Laston, D., Albone, E., Fay, P. J. & Bambara, R. A. Characterization of an RNase H deficient mutant of human immunodeficiency virus-1 reverse transcriptase having an aspartate to asparagine change at position 498. *Biochim. Biophys. Acta - Gene Structure & Expression* **1219**, 380-388 (1994).

194. Mizrahi, V., Usdin, M. T., Harington, A. & Dudding, L. R. Site-directed mutatgenesis of the conserved Asp-443 and Asp-498 carboxy-terminal residues of HIV-1 reverse transcriptase. *Nucleic Acids Res.* **18**, 5359-5363 (1990).

195. Smith, C. M., Smith, J. S. & Roth, M. J. RNase H requirements for the second strand transfer reaction of human immunodeficiency virus type 1 reverse transcription. *J. Virol.*. **73**, 6573-6581 (1999).

196. Furfine, E. S. & Reardon, J. E. Human immunodeficiency virus reverse transcriptase ribonuclease H: Specificity of $tRNA^{Lys3}$-primer excision. *Biochemistry* **30**, 7041-7046 (1991).

197. Furfine, E. S. & Reardon, J. E. Reverse transcriptase RNase H from the human immunodeficiency virus: Relationship of the DNA polymerase and RNA hydrolysis activities. *J. Biol. Chem.* **266**, 406-412 (1991).

198. Schatz, O., Mous, J. & Legrice, S. F. J. HIV-1 RT-associated ribonuclease H displays both endonuclease and 3'-5' exonuclease activity. *EMBO Journal*, 4 (1990).

199. Gopalakrishnan, V., Peliska, J. A. & Benkovic, S. J. Human immunodeficiency virus type 1 reverse transcriptase: Spatial and temporal relationship between the polymerase and RNase H activities. *Proceed. Natl. Acad. Sci. USA.* **89**, 10763-10767 (1992).

200. Artzi, H. B., Shemesh, J., Zeelon, E., Amit, B., Kleiman, L., Gorecki, M. & Panet, A. Ribonuclease H activity during initiation of reverse transcription using $tRNA^{Lys}$/RNA primer/template of human immunodeficiency virus. *Arch. Biochem. Biophys.* **325**, 209-216 (1996).

201. Smith, J. S. & Roth, M. J. Specificity of human immunodeficiency virus-1 reverse rranscriptase-associated ribonuclease H in removal of the minus-strand primer, $tRNA^{Lys3}$. *J. Biol. Chem.* **267**, 15071-15079 (1992).

202. Wohrl, B. & Moelling, K. Interaction of HIV-1 ribonuclease H with polypurine tract containing RNA-DNA hybrids. *Biochemistry* **29**, 10141-10147 (1990).

203. Zhan, X. Y., Tan, C. K., Scott, W. A., Mian, A. M., Downey, K. M. & So, A. G. Catalytically distinct conformations of the ribonuclease H of HIV-1 reverse transcriptase by substrate cleavage patterns and inhibition by azidothymidylate and *N*-ethylmaleimide. *Biochemistry* **33**, 1366-1372 (1994).

204. Cowan, J. A., Ohyama, T., Howard, K., Rausch, J. W., Cowan, S. M. L. & Le Grice, S. F. J. Metal-ion stoichiometry of the HIV-1 RT ribonuclease H domain: evidence for two mutually exclusive sites leads to new mechanistic insights on metal-mediated hydrolysis in nucleic acid biochemistry. *J. Biol. Inorg. Chem.* **5**, 67-74 (2000).

205. Ohtani, N., Haruki, M., Morikawa, M. & Kanaya, S. Molecular diversities of RNases H. *J. Biosci. Bioeng.* **88**, 12-19 (1999).

206. Ding, J. P., Hughes, S. H. & Arnold, E. Protein-nucleic acid interactions and DNA conformation in a complex of human immunodeficiency virus type 1 reverse transcriptase with a double-stranded DNA template-primer. *Biopolymers* **44**, 125-138 (1997).

207. Fan, N. S., Rank, K. B., Slade, D. E., Poppe, S. M., Evans, D. B., Kopta, L. A., Olmsted, R. A., Thomas, R. C., Tarpley, W. G. & Sharma, S. K. A drug resistance mutation in the inhibitor binding pocket of human immunodeficiency virus type 1 reverse transcriptase impairs DNA synthesis and RNA degradation. *Biochemistry* **35**, 9737-9745 (1996).

208. Gao, H. Q., Boyer, P. L., Arnold, E. & Hughes, S. H. Effects of mutations in the polymerase domain on the polymerase, RNase H and strand transfer activities of human immunodeficiency virus type 1 reverse transcriptase. *J. Mol. Biol.* **277**, 559-572 (1998).

209. Palaniappan, C., Fay, F. J. & Bambara, R. A. Nevirapine alters the cleavage specificity of ribonuclease H of human immunodeficiency virus 1 reverse transcriptase. *J. Biol. Chem.* **270**, 4861-4869 (1995).

210. Isel, C., Keith, G., Ehresmann, B., Ehresmann, C. & Marquet, R. Mutational analysis of the $tRNA^{Lys3}$/HIV-1 RNA (primer/template) complex. *Nucleic Acids Res.* **26**, 1198-1204 (1998).

211. Isel, C., Westhof, E., Massire, C., Le Grice, S. F. J., Ehresmann, B., Ehresmann, C. & Marquet, R. Structural basis for the specificity of the initiation of HIV-1 reverse transcription. *EMBO Journal* **18**, 1038-1048 (1999).

212. Wohrl, B. M., Krebs, R., Goody, R. S. & Restle, T. Refined model for primer/template binding by HIV-1 reverse transcriptase: Pre-steady-state kinetic analyses of primer/template binding and nucleotide incorporation events distinguish between different binding modes depending on the nature of the nucleic acid substrate. *J. Mol. Biol.*. **292**, 333-344 (1999).

213. Yusupova, G., Lanchy, J. M., Yusupov, M., Keith, G., Legrice, S. F. J., Ehresmann, C., Ehresmann, B. & Marquet, R. Primer selection by HIV-1 reverse transcriptase on RNA-$tRNA^{Lys3}$ and DNA-$tRNA^{Lys3}$ hybrids. *J. Mol. Biol.* **261**, 315-321 (1996).

214. Smith, R. H., Michejda, C. J., Hughes, S. H., Arnold, E., Janssen, P. A. J. & Smith, M. B. K. Structure and mechanism of action of nonnucleoside inhibitors of HIV-1 reverse transcriptase - Strategies to combat drug resistance. *Theochem-J. Mol. Struc.* **423**, 67-77 (1998).

215. Hsiou, Y., Das, K. Y., Ding, J. P., Clark, A. D., Kleim, J. P., Rosner, M., Winkler, I., Riess, G., Hughes, S. H. & Arnold, E. Structures of Tyr188Leu mutant and wild-type HIV-1 reverse transcriptase complexed with the non-nucleoside inhibitor HBY 097 - Inhibitor flexibility is a useful design feature for reducing drug resistance. *J. Mol. Biol.* **284**, 313-323 (1998).

216. Hopkins, A. L., Ren, J. S., Esnouf, R. M., Willcox, B. E., Jones, E. Y., Ross, C., Miyasaka, T., Walker, R. T., Tanaka, H., Stammers, D. K. & Stuart, D. I. Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT series reveal conformational changes relevant to the design of potent non-nucleoside inhibitors. *J. Med. Chem.* **39**, 1589-1600 (1996).

217. Esnouf, R. M., Ren, J. S., Hopkins, A. L., Ross, C. K., Jones, E. Y., Stammers, D. K. & Stuart, D. I. Unique features in the structure of the complex between HIV-1 reverse transcriptase and the bis(heteroaryl)piperazine (BHAP) U-90152 explain resistance mutations for this nonnucleoside inhibitor. *Proceed. Natl. Acad. Sci. USA.* **94**, 3984-3989 (1997).

218. Ren, J. S., Esnouf, R. M., Hopkins, A. L., Jones, E. Y., Kirby, I., Keeling, J., Ross, C. K., Larder, B. A., Stuart, D. I. & Stammers, D. K. 3'-Azido-3'-deoxythymidine drug resistance mutations in HIV-1 reverse transcriptase can induce long range conformational changes. *Proceed. Natl. Acad. Sci. USA.* **95**, 9518-9523 (1998).

219. Smerdon, S. J., Jäger, J., Wang, J., Kohlstaedt, L. A., Chirino, A. J., Friedman, J. M., Rice, P. A. & Steitz, T. A. Structure of the binding site for nonnucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. *Proceed. Natl. Acad. Sci. USA.* **91**, 3911-3915 (1994).

220. Maga, G., Amacker, M., Ruel, N., Hubscher, U. & Spadari, S. Resistance to nevirapine of HIV-1 reverse transcriptase mutants - Loss of stabilizing interactions and thermodynamic or steric barriers are induced by different single amino acid substitutions. *J. Mol. Biol.* **274**, 738-747 (1997).

221. Bacheler, L. T. Resistance to non-nucleoside inhibitors of HIV-I reverse transcriptase. *Drug Resistance Updates.* **2**, 56-67 (1999).

222. Hsiou, Y., Ding, J., Das, K., Clark, A. D., Hughes, S. H. & Arnold, E. Structure of unliganded HIV-1 reverse transcriptase at 2.7 Å resolution - Implications of conformational changes for polymerization and inhibition mechanisms. *Structure* **4**, 853-860 (1996).

223. Ding, J. P., Das, K., Hsiou, Y., Zhang, W., Arnold, E., Yadav, P. N. S. & Hughes, S. H. "Structural studies of HIV-1 reverse transcriptase and implications for drug design." in *Structure-Based Drug Design* 41-82 (Marcel Dekker, New York, 1997).

224. Doublie, S., Sawaya, M. R. & Ellenberger, T. An open and closed case for all polymerases. *Structure* **7**, R31-R35 (1999).

225. Smith, M. B. K., Michejda, C. J., Hughes, S. H., Boyer, P. L., Janssen, P. A. J., Andries, K., Buckheit, R. W. & Smith, R. H. Molecular modeling of HIV-1 reverse transcriptase drug-resistant mutant strains - Implications for the mechanism of polymerase action. *Prot. Eng.* **10**, 1379-1383 (1997).

226. Fedoroff, O. Y., Salazar, M. & Reid, B. R. Structure of a DNA-RNA hybrid duplex - Why RNase H does not cleave pure RNA. *J. Mol. Biol.* **233**, 509-523 (1993).

227. Isel, C., Ehresmann, C., Keith, G., Ehresmann, B. & Marquet, R. Initiation of reverse transcription of HIV-1 - Secondary structure of the HIV-1 RNA/$tRNA^{Lys3}$ (template/primer) complex. *J. Mol. Biol.* **247**, 236-250 (1995).

228. Suo, Z. C. & Johnson, K. A. Effect of RNA secondary structure on the kinetics of DNA synthesis catalyzed by HIV-1 reverse transcriptase. *Biochemistry* **36**, 12459-12467 (1997).

229. Lu, X. J., Shakked, Z. & Olson, W. K. A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* **300**, 819-840 (2000).

230. Barat, C., Lullien, V., Schatz, O., Keith, G., Nugeyre, M. T., Gruninger-Leitch, F., Barre-Sinoussi, F., LeGrice, S. F. & Darlix, J. L. HIV-1 reverse transcriptase specifically interacts with the anticodon domain of its cognate primer tRNA. *EMBO Journal* **8**, 3279-3285 (1989).

231. Powell, M. D. & Levin, J. G. Sequence and structural determinants required for priming of plus-strand DNA synthesis by the human immunodeficiency virus type 1 polypurine tract. *J. Virol.* **70**, 5288-5296 (1996).

232. Huber, H. E., McCoy, J. M., Seehra, J. S. & Richardson, C. C. Human immunodeficiency virus 1 reverse transcriptase: Template binding, processivity, strand displacement synthesis, and template switching. *J. Biol. Chem.* **264**, 4669-4678 (1989).

233. Reardon, J. E. Human immunodeficiency virus reverse transcriptase: A kinetic analysis of RNA-dependent and DNA-dependent polymerization. *J. Biol. Chem.* **268**, 8743-8751 (1993).

234. Lanchy, J. M., Ehresmann, C., Legrice, S. F. J., Ehresmann, B. & Marquet, R. Binding and kinetic properties of HIV-1 reverse transcriptase markedly differ during initiation and elongation of reverse transcription. *EMBO Journal* **15**, 7178-7187 (1996).

235. Thrall, S. H., Krebs, R., Wohrl, B. M., Cellai, L., Goody, R. S. & Restle, T. Pre-steady-state kinetic characterization of RNA-primed initiation of transcription by HIV-1 reverse transcriptase and analysis of the transition to a processive DNA-primed polymerization mode. *Biochemistry* **37**, 13349-13358 (1998).

236. Boyer, J. C., Bebenek, K. & Kunkel, T. A. Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. *Proceed. Natl. Acad. Sci. USA.* **89**, 6919-6923 (1992).

237. Essink, B. B. O. & Berkhout, B. The fidelity of reverse transcription differs in reactions primed with RNA versus DNA primers. *J. Biomed. Sci.* **6**, 121-132 (1999).

238. Destefano, J. J., Mallaber, L. M., Fay, P. J. & Bambara, R. A. Determinants of the RNase H cleavage specificity of human immunodeficiency virus reverse transcriptase. *Nucleic Acids Res.* **21**, 4330-4338 (1993).

239. Destefano, J. J., Mallaber, L. M., Fay, P. J. & Bambara, R. A. Quantitative analysis of RNA cleavage during RNA-directed DNA synthesis by human immunodeficiency and avian myeloblastosis virus reverse transcriptases. *Nucleic Acids Res.* **22**, 3793-3800 (1994).

240. Gotte, M., Fackler, S., Hermann, T., Perola, E., Cellai, L., Gross, H. J., Legrice, S. F. J. & Heumann, H. HIV-1 reverse transcriptase-associated RNase H cleaves RNA/RNA in arrested complexes - Implications for the mechanism by which RNase H discriminates between RNA/RNA and RNA/DNA. *EMBO Journal* **14**, 833-841 (1995).

241. Reddy, M. R. & Parrill, A. L. "Overview of rational drug design." in *Rational Drug Design: Novel Methodology and Practical Applications* (eds. Reddy, M. R. & Parrill, A. L.) 1-11 (American Chemical Society, Washington, 1999).

242. Lunney, E. A. & Humblet, C. "Structure-based design: From renin to HIV-1 protease." in *Structure-Based Ligand Design* (eds. Gubernator, K. & Boehm, H. J.) 37-71 (Wiley-VCH, Weinheim, 1998).

243. Schafer, W. "Rational design of inhibitors of HIV-1 reverse transcriptase." in *Structure-Based Ligand Design* (eds. Gubernator, K. & Boehm, H. J.) 122-127 (Wiley-VCH, Weinheim, 1998).

244. Gubernator, K., Heinze-Krauss, I., Angehrn, P., Charnas, R. L., Hubschwerlen, C., Oefner, C., Page, M. G. P. & Winkler, F. K. "Structure-based design of potent beta-lactamase inhibitors." in *Structure-Based Ligand Design* (eds. Gubernator, K. & Boehm, H. J.) 89-103 (Wiley-VCH, Weinheim, 1998).

245. Taylor, N. R. "Inhibition of sialidase." in *Structure-Based Ligand Design* (eds. Gubernator, K. & Boehm, H. J.) 105-119 (Wiley-VCH, Weinheim, 1998).

246. Sprague, P. W. Automated chemical hypothesis generation and database searching with Catalyst®. *Perspectives in Drug Discovery & Design* **3**, 1-20 (1995).

247. Clark, D. E., Murray, C. W. & Li, J. "Current issues in *de novo* molecular design." in *Reviews in Computational Chemistry* (eds. Lipkowitz, K. B. & Boyd, D. B.) 67-125 (Wiley-VCH, New York, 1997).

248. Gubernator, K. & Boehm, H. J. "Examples of active areas of structure based-design." in *Structure-Based Ligand Design* (eds. Gubernator, K. & Boehm, H. J.) 15-36 (Wiley-VCH, Weinheim, 1998).

249. Gubernator, K. & Boehm, H. J. "Rational design of bioactive molecules." in *Structure-Based Ligand Design* (eds. Gubernator, K. & Boehm, H. J.) 1-13 (Wiley-VCH, Weinheim, 1998).

250. Sprague, P. W. & Hoffmann, R. "CATALYST pharmacophore models and their utility as queries for searching 3D databases." in *Computer-Assisted Lead Finding and Optimization: Current Tools for Medicinal Chemistry* (eds. van de Waterbeemd, H., Testa, B. & Folkers, G.) 225-240 (VHCA Wiley-VCH, Weinheim, 1997).

251. Kubinyi, H. "Comparative molecular field analysis (COMFA)." in *Encyclopedia of Computational Chemistry* (eds. Schleyer, P. v. R. *et al.*) 448-460 (John Wiley & Sons Ltd, Chichester, UK, 1998).

252. Kubinyi, H. "Quantitative structure-activity relationships in drug design." in *Encyclopedia of Computational Chemistry* (eds. Schleyer, P. v. R. *et al.*) 2309-2320 (John Wiley & Sons Ltd, Chichester, UK, 1998).

253. Greco, G., Novellino, E. & Martin, Y. C. "Approaches to three-dimensional quantitative structure-activity relationships." in *Reviews in Computational Chemistry* (eds. Lipkowitz, K. B. & Boyd, D. B.) 183-240 (Wiley-VCH, New York, 1997).

254. Oprea, T. I. & Waller, C. L. "Theoretical and practical aspects of three-dimemensional quantitative structure-activity relationships." in *Reviews in Computational Chemistry* (eds. Lipkowitz, K. B. & Boyd, D. B.) 127-182 (Wiley-VCH, New York, 1997).

255. Hogberg, T. & Norinder, U. "Theoretical and experimental methods in drug design applied on antipsychotic dopamine antagonists." in *A Textbook of Drug Design and Development* (eds. Krogsgaard-Larsen, P. & Bundgaard, H.) 55-91 (Harwood Academic Publishers, Chur, Switzerland, 1991).

256. Cramer, R. D. I., Patterson, D. E. & Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **110**, 5959-5967 (1988).

257. Barreca, M. L., Carotti, A., Carrieri, A., Chimirri, A., Monforte, A. M., Calace, M. P. & Rao, A. Comparative molecular field analysis (CoMFA) and docking studies of non-nucleoside HIV-1 RT inhibitors (NNIs). *Bioorg. Med. Chem.* **7**, 2283-2292 (1999).

258. Clark, R. D., Leonard, J. M. & Strizhev, A. "Pharmacophore models and comparative molecular field analysis (CoMFA)." in *Pharmacophore Perception, Development and Use in Drug Design* (ed. Guner, O. F.) 153-169 (International University Line, La Jolla, California, 2000).

259. Diana, G. D., Kowalczyk, P., Treasurywala, A. M., Oglesby, R. C., Pevear, D. C. & Dutko, F. J. CoMFA analysis of the interaction of antipicornavirus compounds in the binding pocket of human rhinovirus-14. *J. Med. Chem.* **35**, 1002-1008 (1992).

260. Klebe, G. & Abraham, U. On the prediction of binding properties of drug molecules by comparative molecular field analysis. *J. Med. Chem.* **35**, 70-80 (1993).

261. Depriest, S. A., Mayer, D., Naylor, C. B. & Marshall, G. R. 3D-QSAR of angiotensin-converting enzyme and thermolysin inhibitors - A comparison of CoMFA models based on deduced and experimentally determined active site geometries. *J. Am. Chem. Soc.* **115**, 5372-5384 (1993).

262. Klebe, G., Abraham, U. & Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **37**, 4130-4146 (1994).

263. Silverman, B. D. & Platt, D. E. Comparative molecular moment analysis (CoMMA) - 3D-QSAR without molecular superposition. *J. Med. Chem.* **39**, 2129-2140 (1996).

264. Milne, G. W. A. "Pharmacophore and drug discovery." in *Encyclopedia of Computational Chemistry* (eds. Schleyer, P. v. R. *et al.*) 2046-2056 (John Wiley & Sons Ltd, Chichester, UK, 1998).

265. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **28** (2000).

266. Beddell, C. R. Designing drugs to fit a macromolecular receptor. *Chem. Soc. Rev.* **13**, 279-319 (1984).

267. Blaney, J. m. & Dixon, J. S. A good ligand is hard to find: Automated docking methods. *Perspectives in Drug Discovery & Design* **1**, 301-319 (1993).

268. Joseph-McCarthy, D. Computational approaches to structure-based ligand design. *Pharmacol. Ther.*. **84**, 179-191 (1999).

269. Koehler, K. F., Rao, S. N. & Snyder, J. P. "Modelling drug-receptor interactions." in *Guidebook on Molecular Modelling in Drug Design* (ed. Cohen, N. C.) 235-336 (Academic Press, San Diego, 1996).

270. Jones, G. & Willet, P. Docking small-molecule ligands into active sites. *Curr. Opin. Biotech.* **213**, 627-630 (1995).

271. Rosenfeld, R., Vajda, S. & Delisi, C. Flexible docking and design. *Ann. Rev. Biophys. Biomol. Struct.* **24**, 677-700 (1995).

272. Oshiro, C. M., Kuntz, I. D. & Dixon, J. S. Flexible ligand docking using a genetic algorithm. *J. Comp. Aided. Mol. Des.* **9**, 113-130 (1995).

273. Lybrand, T. P. Ligand protein docking and rational drug design. *Curr. Opin. Struct. Biol.* **5**, 224-228 (1995).

274. Leach, A. R. & Kuntz, I. D. Conformational analysis of flexible ligands in macromolecular receptor sites. *J. Comp. Chem.* **13**, 730-748 (1992).

275. Leach, A. R. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* **235**, 345-356 (1994).

276. Goodford, P. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28**, 849-857 (1985).

277. Bohm, H. J. LUDI: Rule-based automatic design of new substituents for enzyme inhibitor leads. *J. Comp. Aided. Mol. Des.* **6**, 593-606 (1992).

278. Bohm, H. J. Site-directed structure generation by fragment-joining. *Perspectives in Drug Discovery & Design* **3**, 21-33 (1995).

279. Bohm, H. J. & Fisccher, S. "*De novo* ligand design." in *Encyclopedia of Computational Chemistry* (eds. Schleyer, P. v. R. *et al.*) 657-663 (John Wiley & Sons Ltd, Chichester, UK, 1998).

280. Unity 3D® 4.1 (Tripos, St. Louis, Missouri, 1999).

281. Cerius^2 Structure Based Focusing® 4.5 (Molecular Simulations Inc., San Diego, CA, USA, 2000).

282. Venkatachalam, C. M., Kirchhoff, P. & Waldman, M. "Receptor-based pharmacophore perception and modeling." in *Pharmacophore Perception, Development and Use in Drug Design* (ed. Guner, O. F.) 342-350 (International University Line, La Jolla, California, 2000).

283. Vedani, A., Dobler, M. & Zbinden, P. Quasi-atomistic receptor surface models - A bridge between 3-D QSAR and receptor modeling. *J. Am. Chem. Soc.* **120**, 4471-4477 (1998).

284. Griffith, R., Bremner, J. B. & Coban, B. "Docking-derived pharmacophores from models of receptor-ligand complexes." in *Pharmacophore Perception, Development and Use in Drug Design* (ed. Guner, O. F.) 387-408 (International University Line, La Jolla, California, 2000).

285. Bremner, J. B., Coban, B., Griffith, R., Groenewoud, K. M. & Yates, B. F. Ligand design for α_1 adrenoceptor subtype selective antagonists. *Bioorg. Med. Chem.* **8**, 201-214 (2000).

286. Thomas IV, B. E., Joseph-McCarthy, D. & Alvarez, J. C. "Pharmacophore-based molecular docking." in *Pharmacophore Perception, Development and Use in Drug Design* (ed. Guner, O. F.) 353-367 (International University Line, La Jolla, California, 2000).

287. Schleifer, K. J., Tot, E. & Holtje, H. D. Pharmacophore and pseudoreceptor modelling of class Ib antiarrhythmic and local anaesthetic lidocaine analogues. *Pharmazie* **53**, 596-602 (1998).

288. Zbinden, P., Dobler, M., Folkers, G. & Vedani, A. PrGen - Pseudoreceptor modeling using receptor-mediated ligand alignment and pharmacophore equilibration. *Quantit. Struct. Act. Relat.* **17**, 122-130 (1998).

289. Greenidge, P. A., Carlsson, B., Bladh, L. G. & Gillner, M. Pharmacophores incorporating numerous excluded volumes defined by x-ray crystallographic structure in three-dimensional database searching - Application to the thyroid hormone receptor. *J. Med. Chem.* **41**, 2503-2512 (1998).

290. Gillner, M. & Greenidge, P. "The use of multiple excluded volumes derived from x-ray crystallographic structures in 3D database searching and 3D QSAR." in *Pharmacophore Perception, Development and Use in Drug Design* (ed. Guner, O. F.) 373-384 (International University Line, La Jolla, California, 2000).

291. Catalyst® 4.0 (Molecular Simulations Inc., San Diego, CA, USA, 1998).

292. LeapFrog® 4.1 (Tripos, St. Louis, Missouri, 1999).

293. Cerius^2 LUDI® 4.5 (Molecular Simulations Inc., San Diego, CA, USA, 2000).

294. Garg, R., Gupta, S. P., Gao, H., Babu, M. S., Debnath, A. K. & Hansch, C. Comparative quantitative structure-activity relationship studies on anti-HIV drugs. *Chem. Rev.* **99**, 3525-3601 (1999).

295. Clement, O. O. & Mehl, A. T. "HipHop: Pharmacophores based on multiple common feature alignments." in *Pharmacophore Perception, Development and Use in Drug Design* (ed. Guner, O. F.) 71-84 (International University Line, La Jolla, California, 2000).

296. Li, H., Sutter, J. & Hoffmann, R. "HypoGen: An automated systems for generating 3D predictive pharmacophore models." in *Pharmacophore Perception, Development and Use in Drug Design* (ed. Guner, O. F.) 173-189 (International University Line, La Jolla, California, 2000).

297. Catalyst notes 4.0 (Molecular Simulations Inc., 1998). http://www.msi.com/support/catalyst.

298. Catalyst tutorials 4.0 (Molecular Simulations Inc., 1998). http://www.msi.com/doc/catalyst40/tutorials.

299. Sober, E. "Simplicity." in *A companion to the philosophy of science* (ed. Newton-Smith, W. H.) (Blackwell Publishers Ltd, Oxford, UK, 2000).

300. Greene, J., Kahn, S., Savoj, H., Sprague, P. & Teig, S. Chemical function queries for 3D database search. *J. Chem. Inform. Comp. Sci.* **34**, 1297-1308 (1994).

301. Archive of catalyst mailing list (Molecular Simulations Inc., 2000). http://www.msi.com/user/groups/catalyst/cat-archive-dir/index.html.

302. Catalyst 4.6® (Molecular Simulations Inc., San Diego, CA, USA, 2000).

303. Ren, J. S., Esnouf, R. M., Hopkins, A. L., Stuart, D. I. & Stammers, D. K. Crystallographic analysis of the binding modes of thiazoloisoindolinone non-nucleoside inhibitors to HIV-1 reverse transcriptase and comparison with modeling studies. *J. Med. Chem.* **42**, 3845-3851 (1999).

304. Ren, J. S., Nichols, C., Bird, L. E., Fujiwara, T., Sugimoto, H., Stuart, D. I. & Stammers, D. K. Binding of the second generation non-nucleoside inhibitor S-1153 to HIV-1 reverse transcriptase involves extensive main chain hydrogen bonding. *J. Biol. Chem.* **275**, 14316-14320 (2000).

305. Fujiwara, T., Sato, A., Elfarrash, M., Miki, S., Abe, K., Isaka, Y., Kodama, M., Wu, Y. M., Chen, L. B., Harada, H., Sugimoto, H., Hatanaka, M. & Hinuma, Y. S-1153 inhibits replication of known drug-resistant strains of human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* **42**, 1340-1345 (1998).

306. Mertens, A., Zilch, H., Konig, B., Schafer, W., Poll, T., Kampe, W., Seidel, H., Lesser, U. & Leinert, H. Selective non-nucleoside HIV-1 reverse transcriptase inhibitors. New 2,3-dihydrothiazolo[2,3-a]isoindol-5(9b*H*)-ones and related compounds with anti-HIV activity. *J. Med. Chem.* **36**, 2526-2535 (1993).

307. Catalyst® 4.5 (Molecular Simulations Inc., San Diego, CA, USA, 1999).

308. SiteID® 4.1 (Tripos, St. Louis, Missouri, 1999).

309. Sybyl® 6.6 (Tripos, St. Louis, Missouri, 1999).

310. SiteID Design Manual 6.6 (Tripos, 1999). http://www.tripos.com.

311. Carlson, H. A. & McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol. Pharm.* **57**, 213-218 (2000).

312. Sudbeck, E. A., Mao, C., Vig, R., Venkatachalam, T. K., Tuel-Ahlgren, L. & Uckun, F. M. Structure-based design of novel dihydroalkoxybenzyloxopyrimidine derivatives as potent nonnucleoside inhibitors of the human immunodeficiency virus reverse transcriptase. *Antimicrob. Agents Chemother.* **42**, 3225-3233 (1998).

313. Carlson, H. A., Masukawa, K. M. & McCammon, J. A. Method for including the dynamic fluctuations of a protein in computer-aided drug design. *J. Phys. Chem.* **103**, 10213-10219 (1999).

314. Masukawa, K. M., Carlson, H. A. & McCammon, J. A. "Technique for developing a pharmacophore model that accommodates inherent protein flexibility: An application to HIV-1 integrase." in *Pharmacophore Perception, Development and Use in Drug Design* (ed. Guner, O. F.) 412-427 (International University Line, La Jolla, California, 2000).

315. Carlson, H. A., Masukawa, K. M., Rubins, K., Bushman, F. D., Jorgensen, W. L., Lins, R. D., Briggs, J. M. & McCammon, J. A. Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* **43**, 2100-2114 (2000).

316. Marriott, D. P., Dougall, I. G., Meghani, P., Liu, Y. J. & Flower, D. R. Lead generation using pharmacophore mapping and three-dimensional database searching: Application to muscarinic M_3 receptor antagonists. *J. Med. Chem.* **42**, 3210-3216 (1999).

317. Fox, T. & Haaksma, E. E. J. Computer based screening of compound databases: 1. Preselection of benzamidine-based thrombin inhibitors. *J. Comp. Aided. Mol. Des.* **14**, 411-425 (2000).

318. Unity User Guide 4.1 (Tripos, 1999). http://www.tripos.com.

319. Development Therapeutics Program (National Cancer Institute, 2000). http://dtp.nci.nih.gov/main.html and http://dtp.nci.nih.gov/docs/dtp_search.html.

320. Protein Data Bank (Research Collaboratory for Structural Bioinformatics, 2000). http://www.rcsb.org.

321. Genin, M. J., Poel, T. J., Yagi, Y., Biles, C., Althaus, I., Keiser, B. J., Kopta, L. A., Friis, J. M., Reusser, F., Adams, W. J., Olmsted, R. A., Voorman, R. L., Thomas, R. C. & Romero, D. L. Synthesis and bioactivity of novel bis(heteroaryl)piperazine (BHAP) reverse transcriptase inhibitors - Structure-activity relationships and increased metabolic stability of novel substituted pyridine analogs. *J. Med. Chem.* **39**, 5267-5275 (1996).

322. Baba, M., Balzarini, J., Pauwels, R. & De Clercq, E. "HIV-1 specific reverse transcriptase inhibitors." in *Anti-AIDS Drug Development Challenges, Strategies and Prospects* (eds. Mohan, P. & Baba, M.) 239-267 (Harwood Academic, 1995).

323. Pauwels, R., Andries, K., Desmyter, J., Schols, D., Kukla, M. J., Breslin, H. J., Raeymaeckers, A., Gelder, J. V., Woestenborghs, R., Heykants, J., Schellekens, K., Janssen, M. A. C., De Clerq, E. & Janssen, P. A. J. Potent and selective inhibition of HIV-1 replication *in vitro* by a novel series of TIBO derivatives. *Nature* **343**, 470-474 (1990).

324. Balzarini, J., Perezperez, M. J., Velazquez, S., Sanfelix, A., Camarasa, M. J., Declercq, E. & Karlsson, A. Suppression of the breakthrough of human immunodeficiency virus type 1 (HIV-1) in cell culture by thiocarboxanilide derivatives when used individually or in combination with other HIV-1 specific Inhibitors (ie, TSAO derivatives). *Proceed. Natl. Acad. Sci. USA.* **92**, 5470-5474 (1995).

325. McMahon, J. B., Buckheit, R. W., Gulakowski, R. J., Currens, M. J., Vistica, D. T., Shoemaker, R. H., Stinson, S. F., Russell, J. D., Bader, J. P., Narayanan, V. L., Schultz, R. J., Brouwer, W. G., Felauer, E. E. & Boyd, M. R. Biological and biochemical anti-human immunodeficiency virus activity of UC 38, a new non-nucleoside reverse transcriptase inhibitor. *J Pharmacol Exp Thers* **276**, 298-305 (1996).

326. Hargrave, K. D., Proudfoot, J. R., Grozinger, K. G., Cullen, E., Kapadia, S. R., Patel, U. R., Fuchs, V. U., Mauldin, S. C., Vitous, J., Behnke, M. L., Klunder, J. M., Pal, K., Skiles, J. W., McNeil, D. W., Rose, J. M., Chow, G. C., Skoog, M. T., Wu, J. C., Schmidt, G., Engel, W. W., Eberlein, W. G., Saboe, T. D., Campbell, S. J., Rosenthal, A. S. & Adams, J. Novel non-nucleoside inhibitors of HIV-1 reverse transcriptase. 1. Tricyclic pyridobenzo- and dipyridodiazepinones. *J. Med. Chem.* **34**, 2231-2241 (1991).

327. Miyasaka, T., Tanaka, H., Baba, M., Hayakawa, H., Walker, R. T., Balzarini, J. & De Clerq, E. A novel lead for specific anti-HIV agents: 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine. *J. Med. Chem.* **32**, 2507-2509 (1989).

328. Pauwels, R., Andries, K., Debyser, Z., Van Daele, P., Schols, D., Stoffels, P., De Vreese, K., Woestenborghs, R., Vandamme, A. M., Janssen, C. G. M., Anne, J., Cauwenbergh, G., Desmyter, J., Heykants, J., Janssen, M. A. C., De Clerq, E. & Janssen, P. A. J. Potent and highly selective human immunodeficiency virus type 1 (HIV-1) inhibition by a series of α-anilinophenylacetamide derivatives targeted at HIV-1 reverse transcriptase. *Proc. Natl Acad. Sci. USA* **90**, 1711-1715 (1993).

Appendices

Appendix 7.1 Subdomains of reverse transcriptase

Subdomain	Residues (CABM[124])	Residues (OCMS[133])
fingers	1-84 and 120-150	1-88 and 121-146
	1-85 and 119-149	
palm	85-119 and 151-243	89-120 and 147-242
	86-118 and 150-244	
thumb	244-322	243-311
	245-320	
connection	323-437	312-425
	321-427	
RNase H	438-560	426-560

Subdomains of RT as defined by the CABM (Center for Advanced Biotechnology and Medicine, Rutgers University, New Jersey) and OCMS (Oxford Centre for Molecular Sciences, Oxford) research groups and their associates.[124,133]

Appendix 7.2 Secondary structure of reverse transcriptase

CABM[132]			OCMS[133]	
sequence		secondary structure	sequence	secondary structure
p66	p51			
7-24	7-22	β0-β1	28-43	α1
28-44	28-44	αA	47-75	β1-β3
49-77	49-76	β2-β4	78-83	α2
78-83	78-84	αB	105-110	β4
86-112	87-111	β5a-β6	112-128	α3-α4
114-127	112-127	αC-αD	130-146	β5-β6
128-147	128-147	β7-β8	156-174	α5
155-174	155-174	αE	179-191	β7-β8
178-191	179-191	β9-β10	195-211	α6
195-212	198-212	αF	227-241	β9-β11
214-242		β11a-β14	254-311	α7-α9
	214-219	β11a	326-355	β12-β14
	239-242	β14	364-383	α10
255-311	254-310	αH-αJ	388-391	β15
316-358	316-358	β15-β18	395-404	α11
364-382	364-381	αK	413-416	β16
388-391	386-392	β19	421-428	α12
395-404	395-404	αL	439-470	β17-β19
408-430		β20-β22	474-488	α13
	410-416	β20	492-497	β20
438-470		β1'-β3'	500-527	α14-α15
474-488		αA'	530-535	β21
492-497		β4'		
500-508		αB'		
516-527		αD'		
530-536		β5'		
544-555		αE'		

Secondary structure nomenclature/definition for HIV-1 RT from the CABM and OCMS research groups.[132,133]

In the OCMS secondary structure assignation α3 and α4 are 3_{10}-helices; α13-α15, β9-β11, and β17-β21 are only seen in the p66 subunit; and β12 is only present in the p51 subunit.

Appendix 7.3 X-ray crystallographic structures of reverse transcriptase

PDB file	HIV-1 RT Structure	Crystal Form	Resolution Range (Å)	R-Factor/ Free R-Factor (%)	Data Completeness (%)	IC_{50} (nM) [§]	Reference [‡]
1dlo	unliganded RT	*C*2	8.0-2.7	24.9/33.6	99.5[#]	-	222
1hmv	unliganded RT	*C*2	6.0-3.2	25.4/29.7	85	-	135
1rtj *	unliganded RT	$P2_12_12_1$	25.0-2.35	21.9	89.5	-	29
1rtd	RT/DNA/dNTP	$P2_12_12_1$	12.0-3.2	22.4/29.8	92.2	-	136
2hmi	RT/DNA/Fab	$P3_212$	8.0-2.8	27.1/35.2	74.8	-	132,139
1bqm	RT-HBY 097	*C*2	10.0-3.1	25.8/36.2	99.1	45	215
1bqn	RT-HBY 097 Tyr188Leu mutant	*C*2	10.0-3.3	24.5-36.2	95.1	600	215
1hni	RT/Br-α-APA (R95845)	*C*2	10.0-2.8	25.5/36	78.5	-	134
1hnv	RT/8-Cl TIBO (R86183)	*C*2	10.0-3.0	24.9/35.6	81	4.6	30,144
1klm *	RT/BHAP(U-90152)	$P2_12_12_1$	20.0-2.65	23.7/31.3	86.8	1100/260	217,321,322
1rev *	RT/9-Cl TIBO (R82913)	$P2_12_12_1$	25.0-2.6	22.4	80.7	600/1.5	138,323
1rt1	RT/MKC-422	$P2_12_12_1$	25.0-2.55	19.7	95.5	8	216
1rt2	RT/TNK-651	$P2_12_12_1$	20.0-2.55	20.7	85.3	6	216
1rt3	RT/1051U91 RTMC mutant	$P2_12_12_1$	25.0-3.0	26.4/33.5	83.3	400	218
1rt4	RT/UC781	$P2_12_12_1$	30.0-2.9	23.7/29.5	93.6	-	142
1rt5	RT/UC10	$P2_12_12_1$	30.0-2.9	23.1/29.1	86.1	200	142,324

1rt6	RT/UC38	$P2_12_12_1$	30.0-2.8	23.6/33.5	95.7	150	142,325
1rt7	RT/UC84	$P2_12_12_1$	30.0-3.0	25.8/33.4	87.3	280	142,325
1rth	RT/1051U91	$P2_12_12_1$	25.0-2.2	21.4	81.4	400	133,326
1rti	RT/HEPT	$P2_12_12_1$	25.0-3.0	23.6	86.3	17 000	133,327
1tvr	RT/9-Cl TIBO (R82913)	*C*2	10.0-3.0	25.9	84[#]	33	144
1uwb	RT/8-Cl TIBO (R86183) Tyr181Cys mutant	*C*2	15.0-3.2	27.4/36.0	85	130	144
1vrt	RT/nevirapine	$P2_12_12_1$	25.0-2.2	18.6	87.1	84	133,326
1vru	RT/Cl-α-APA (R90385)	$P2_12_12_1$	25.0-2.4	18.7	86.5	5	133,328
3hvt [†]	RT/nevirapine	*C*2	8.0-2.9	26.6	95.6	84	130,219,326

Crystal structures of HIV-1 RT. (Adapted from[223])

[*] Crystal structures were prepared by soaking out HEPT from a HEPT/RT complex, and then soaking in the new inhibitor, if applicable.

[†] Crystal structure was unsuitable for use in superposition because residues with poor or missing sidechain electron density were treated as alanine residues.

[§] IC_{50} values used were obtained preferably from the same publication that reported the crystallographic data. In cases where this was not possible, IC_{50} were obtained from publications reporting the synthesis of the inhibitors. Where more than one IC_{50} was obtained, the first one was used for graphical analysis.

[‡] First (and second reference for 1hnv.pdb) refers to the crystal structure and may also include the IC_{50} value. Following references are for IC_{50} data.

[#] Data completeness overall for range.

Appendix 7.3.1 Additional-ray crystallographic structures of reverse transcriptase

PDB file	HIV-1 RT Structure	Crystal Form	Resolution Range (Å)	R-Factor/ Free R-Factor (%)	Data Completeness (%)	IC_{50} (nM) [§]	Reference [‡]
1ep4	RT/S-1153	$P2_12_12_1$	30.0-2.5	25.4/33.0	90.9[#]	450	304, 305
1c0t	RT/BM+21.1326	$P2_12_12_1$	20.0-2.7	21.4/27.6	81.2[#]	4100	303, 306
1c0u	RT/BM+50.0934	$P2_12_12_1$	15.0-2.52	23.2/29.8	88.8[#]	220	303, 306

Additional crystal structures of HIV-1 RT.

[§] IC_{50} values used were obtained preferably from the same publication that reported the crystallographic data. In cases where this was not possible, IC_{50} were obtained from publications reporting the synthesis of the inhibitors. Where more than one IC_{50} was obtained, the first one was used for graphical analysis.

[‡] First references refers to the crystal structure, the following references are for IC_{50} data.

[#] Data completeness overall for range.

Appendix 7.4 Displacement data for comparison of superimposition techniques (I)

	DDM derived superimposition		Superimposition of Ding *et al.*[139]	
Residue	unliganded	α-APA	unliganded	α-APA
177	1.49	3.68	1.15	0.92
178	0.90	3.04	0.73	0.57
179	1.19	2.97	0.28	0.38
180	1.20	2.81	0.86	0.53
181	1.02	2.80	0.88	0.74
182	1.24	2.72	0.29	0.46
183	1.27	2.76	0.87	0.40
184	2.05	2.82	0.79	0.85
185	2.31	3.10	1.15	0.92
186	1.56	2.87	0.81	0.30
187	1.43	2.66	0.36	0.28
188	0.94	2.81	0.72	0.56
189	1.43	3.22	0.98	0.93
190	1.46	3.78	0.65	0.51
191	1.31	4.14	0.49	0.56
192	1.36	4.59	0.65	0.72
RMSD	0.997	1.897	0.774	0.638

Displacement of C_α atom of an unliganded and a liganded crystal structure superimposed onto 2hmi using difference distance matrices derived subset and a subset of Ding *et al.*[139] Values for the C_α atom displacement using the Ding *et al.* subset have been generated following the reference procedure.

The DDM derived subset is composed of the backbone residues of 376 residues, a total of 1528 atoms. The Ding *et al.* subset is defined as being comprised of the C_α atoms of the β9-β10 hairpin (residues 177-192), a total of 16 atoms.

Crystal structure	PDB file
unliganded	1dlo
α-APA	1vru
unliganded	2hmi

Appendix 7.5 Displacement data for comparison of superimposition techniques (II)

	DDM derived superimposition				Superimposition of Ren *et al.*[138]			
Residue	1051U91	α-APA	HEPT	Cl-TIBO	1051U91	α-APA	HEPT	Cl-TIBO
Pro95	0.50	0.47	0.57	0.21	0.7	0.7	0.7	0.2
Leu100	0.27	0.40	0.26	0.34	0.3	0.2	0.7	0.6
Lys101	0.46	0.32	0.63	0.34	0.4	0.2	0.5	0.5
Lys103	0.65	0.43	1.50	0.22	0.3	0.4	0.3	0.4
Val106	0.73	1.00	1.50	0.69	0.3	0.7	0.4	0.6
Val179	0.43	0.38	1.69	1.01	0.3	0.3	0.7	0.6
Tyr181	0.44	0.29	1.13	0.67	0.5	0.1	0.5	0.4
Tyr183	0.18	0.40	0.56	0.53	0.5	0.2	0.6	0.5
Tyr188	0.69	0.59	1.10	0.93	0.6	0.2	0.5	0.7
Val189	0.60	0.55	1.32	0.78	0.4	0.3	0.3	0.5
Gly190	0.62	0.74	1.29	0.75	0.3	0.4	0.1	0.5
Pro225	0.92	0.29	1.46	0.33	0.4	0.3	0.4	0.2
Phe227	0.99	0.21	1.33	0.71	0.7	0.4	0.4	0.7
Trp229	1.27	0.22	1.30	0.25	0.9	0.1	0.2	0.3
Leu234	0.86	0.36	1.25	0.37	0.3	0.1	0.6	0.2
His235	0.83	0.36	1.75	0.38	0.3	0.1	1.8	0.2
Pro236	0.69	0.58	2.74	0.48	0.3	0.2	2.7	0.6
Tyr318	0.28	0.24	0.35	0.22	0.5	0.1	0.7	0.1
Glu138	0.51	0.73	0.32	0.30	0.4	1.5	0.8	0.2
No. Atoms	1528	1528	1528	1528	102	109	96	109
RMSD	0.74	0.43	0.81	0.43	0.5	0.4	0.5	0.5

Displacement of C_α atom of liganded crystal structures superimposed onto 1vrt using difference distance matrices derived subset and 'core' subset of Ren *et al.*[138] The number of atoms used in each superimposition and the resultant RMSD are noted. Values for the C_α atom displacement using the 'core' subset have been taken directly from the reference paper.

The 'core' subset is defined as being comprised of the C_α atoms of residues 94-118, 156-215 and 225-243 of the palm subdomain, and 317-319 of the connection subdomain of p66, and residues 137-139 from the fingers domain of p51.

Ligand	PDB file
1051U91	1rth
α-APA	1vru
HEPT	1rti
Cl-TIBO	1rev
nevirapine	1vrt

Appendix 7.6 Subset definitions

Subset	Residues
PAS	110, 183-186
primer grip	227-235
template grip	73-83, 86-93, 148-154
RNase H active site	443, 478, 498, 539, 549
NNIBP	100, 101, 103, 106, 179, 181, 188, 190, 227, 229, 234, 235, 236, 318, 138 (p51)

Residues defining subsets investigated. All residues are from the p66 subunit, unless otherwise stated.

Appendix 7.7 Displacement data for the PAS

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2
Asp110 N	0.60	1.98	0.79	2.58	2.62	2.29	2.86	1.46	2.39	2.22	2.21
Asp110 CA	0.61	2.09	0.75	2.79	2.71	2.53	3.06	1.64	2.63	2.42	2.38
Asp110 C	0.41	2.01	0.80	2.56	2.44	2.18	2.63	1.56	2.48	2.28	2.24
Asp110 O	0.62	2.04	0.82	3.08	2.75	2.21	2.70	1.65	2.72	2.51	2.57
Asp110 CB	0.95	2.03	0.67	3.22	3.10	2.96	3.67	1.94	2.94	2.91	3.08
Asp110 CG	1.22	3.53	0.83	3.67	3.71	3.67	4.56	2.82	2.18	2.75	3.34
Asp110 OD1	1.80	4.75	2.96	3.96	4.22	4.19	5.66	3.33	3.67	2.78	3.74
Asp110 OD2	2.99	4.65	1.54	3.78	3.75	2.85	4.33	3.18	0.73	2.61	3.28
Tyr183 N	0.77	1.29	1.09	2.13	1.94	1.64	2.38	1.17	2.13	1.31	1.07
Tyr183 CA	0.75	1.27	1.16	2.18	2.01	1.80	2.37	1.41	2.24	1.53	1.33
Tyr183 C	1.00	1.38	0.97	2.38	2.14	1.98		1.34	2.30	1.51	1.32
Tyr183 O	1.00	1.48	0.93	2.70	2.50	2.66		1.69	2.69	1.85	1.66
Tyr183 CB	0.59	1.35	1.46	2.42	2.16	2.60	2.84	2.13	2.82	2.11	1.91
Tyr183 CG	0.61	1.41	1.58	2.37	2.23	2.28	2.75	2.76	3.04	2.58	2.14
Tyr183 CD1	1.51	1.82	1.85	1.90	1.84	1.76	2.45	3.05	2.95	2.71	2.23
Tyr183 CD2	0.72	1.50	1.58	2.88	2.72	2.63	3.14	3.29	3.52	3.05	2.31
Tyr183 CE1	1.53	1.80	1.91	1.92	2.00	1.47	2.36	3.71	3.26	3.17	2.46
Tyr183 CE2	0.69	1.56	1.64	2.93	2.89	2.35	3.06	3.86	3.72	3.48	2.55
Tyr183 C2											
Tyr183 OH	0.56	1.43	1.73	2.57	2.78	1.41	2.44	4.63	3.84	3.95	2.84
Met184 N	1.56	1.64	0.96	2.23	1.91	1.58	1.88	1.05	1.99	1.28	0.98
Met184 CA	2.16	2.05	0.88	2.52	2.08	2.22	2.09	1.03	2.10	1.27	1.07
Met184 C	2.27	2.23	0.87	2.83	2.58	2.44	2.70	1.20	2.32	1.40	1.38
Met184 O	2.45	2.42	0.88	3.99	3.84	3.15	3.53	1.57	2.52	1.89	2.09
Met184 CB	2.31	1.99	0.95	3.00	2.12	3.25	2.72	1.47	2.59	1.70	1.43
Met184 CG	2.87	3.02	0.99	3.08	2.32	3.57	2.70	2.81	2.75	3.28	1.76
Met184 SD	6.05	5.40	1.17	3.80	4.47	5.13	4.56	2.03	3.96	2.44	2.23
Met184 CE	7.36	5.77	3.79	6.78	6.99	7.34	7.26	4.94	6.02	5.39	4.91

1rt3	1rt4	1rt5	1rt6	1rt7	1rth	1rti	1tvr	1uwb	1vrt	1vru
1.70	1.84	2.13	1.26	1.79	2.49	1.44	2.61	2.78	1.95	2.55
1.91	2.14	2.38	1.42	1.98	2.73	1.61	2.90	2.95	2.26	2.80
1.53	2.09	2.30	1.39	1.85	2.50	1.47	2.23	2.48	2.16	2.60
1.86	2.12	2.38	1.53	2.02	2.60	1.52	1.90	2.48	2.12	2.64
2.51	2.42	2.65	1.50	2.09	3.29	2.10	3.79	3.61	2.59	3.22
2.50	3.40	3.42	2.19	1.95	3.66	2.60	4.55	4.43	3.07	4.27
3.84	4.01	4.38	2.59	3.95	4.02	2.70	5.02	5.16	3.60	4.91
1.85	3.71	3.29	2.46	0.87	3.66	2.97	4.73	4.48	3.09	4.60
0.77	1.75	1.89	1.34	1.41	1.96	1.14	2.74	2.32	1.82	2.26
0.91	1.84	1.94	1.36	1.58	1.97	1.47	2.69	2.25	1.95	2.33
0.90	1.84	2.00	1.43	1.58	1.95	1.42	2.72	2.31	1.89	2.27
1.29	2.24	2.43	1.81	1.94	2.27	1.79	3.33	2.82	2.39	2.60
1.65	2.52	2.54	1.86	2.20	2.57	2.32	3.51	2.79	2.84	2.99
1.98	2.94	2.89	2.24	2.66	3.07	2.61	3.36	2.51	3.06	3.39
2.35	2.96	2.87	4.57	2.76	3.17	2.88	5.06	4.38	2.93	3.35
2.26	3.46	3.37	0.51	3.13	3.64	2.98	1.77	0.92	3.46	3.96
2.83	3.40	3.27	4.97	3.21	3.68	3.19	4.87	4.16	3.15	3.77
2.53	3.84	3.70	0.73	3.56	4.10	3.16	1.61	0.72	3.65	4.31
3.19	4.17	4.03	3.39	4.00	4.53	3.38	2.73	1.75	3.69	4.61
0.55	1.48	1.65	1.21	1.34	1.64	1.16	2.15	1.88	1.46	1.94
0.77	1.64	1.78	1.28	1.37	1.66	1.52	2.49	2.15	1.86	2.10
1.10	1.91	2.11	1.54	1.49	1.88	1.65	2.75	2.48	2.15	2.31
1.38	2.30	2.63	1.99	1.79	2.39	1.91	2.87	2.49	2.61	2.64
1.12	2.16	2.31	1.65	1.86	2.10	2.06	3.10	2.71	2.47	2.52
1.49	2.35	2.59	3.13	3.46	3.73	2.41	3.26	2.99	2.98	3.14
1.87	4.07	4.27	2.53	2.73	3.36	3.11	5.76	4.78	3.40	3.35
4.91	6.34	6.25	5.66	5.59	6.22	3.81	7.92	7.42	4.97	6.04

Appendix 7.7 Displacement data for the PAS (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2	1rt3	1rt4	1rt5	1rt6	1rt7	1rth	1rti	1tvr	1uwb	1vrt	1vru
Asp185 N	2.23	2.27	0.81	1.82	1.59	1.90	2.37	1.08	2.32	1.17	0.96	1.09	1.80	1.96	1.38	1.40	1.75	1.44	2.61	2.73	1.76	2.08
Asp185 CA	2.18	2.31	0.80	2.17	2.09	2.31	2.88	1.41	2.64	1.47	1.26	1.53	2.09	2.32	1.75	1.71	2.02	1.53	2.86	3.15	2.06	2.32
Asp185 C	1.88	2.10	0.93	2.26	2.23	2.46	3.18	1.64	2.83	1.80	1.68	1.72	2.36	2.63	1.98	1.90	2.35	1.70	3.10	3.33	2.43	2.56
Asp185 O	2.31	2.30	0.94	2.42	2.45	2.71	3.51	1.74	2.90	2.02	2.01	1.91	2.66	2.93	2.20	2.03	2.56	2.10	3.14	3.61	2.87	2.91
Asp185 CB	2.29	2.43	0.59	2.20	2.23	2.40	3.28	1.20	2.62	1.23	0.86	1.58	2.13	2.38	1.57	1.60	2.06	1.09	3.04	3.49	1.60	1.90
Asp185 CG	2.26	2.62	0.62	3.31	3.36	3.58	3.34	2.63	4.63	2.41	1.90	3.04	2.09	2.34	3.23	2.95	2.08	0.98	2.97	3.62	3.54	3.31
Asp185 OD1	2.57	3.56	0.84	4.75	4.94	5.88	3.93	5.10	6.59	4.86	4.33	4.96	2.30	2.87	5.68	5.37	2.77	1.09	3.00	5.13	5.77	5.50
Aps185 OD2	2.30	2.38	1.72	3.36	3.39	3.20	3.01	1.46	4.80	1.13	0.68	2.78	1.85	1.84	2.17	1.73	1.52	2.55	3.14	2.55	3.20	2.62
Asp186 N	1.21	1.76	1.04	2.27	2.15	2.39	3.11	1.74	2.95	1.84	1.72	1.72	2.27	2.55	1.96	1.89	2.41	1.54	3.09	3.20	2.35	2.57
Asp186 CA	0.83	1.56	1.20	2.39	2.29	2.52	3.40	1.98	3.09	2.16	2.16	1.95	2.52	2.80	2.17	2.06	2.70	1.64	3.29	3.40	2.71	2.88
Asp186 C	0.68	1.39	0.95	2.38	2.17	2.35	3.04	1.68	2.83	1.86	1.78	1.87	2.23	2.46	1.82	1.71	2.37	1.39	2.95	2.92	2.41	2.62
Asp186 O	0.67	1.38	0.88	3.13	2.37	2.56	3.28	2.18	3.25	2.39	2.33	2.53	2.64	2.78	2.05	1.98	2.39	1.76	3.71	3.19	3.08	3.20
Asp186 CB	0.94	1.39	1.49	2.39	2.44	2.63	3.75	2.16	3.26	2.42	2.56	2.05	2.80	3.09	2.38	2.36	3.04	1.87	3.51	3.78	3.12	3.36
Asp186 CG	1.08	2.73	1.60	2.39	2.57	2.67	4.24	2.44	3.61	2.79	2.86	2.31	3.22	3.56	2.92	2.93	3.38	2.26	3.95	4.44	3.56	3.75
Asp186 OD1	0.97	4.04	1.71	3.19	3.44	3.30	4.86	2.36	3.90	3.11	3.36	2.94	3.68	4.11	3.18	2.80	3.84	2.27	4.82	5.37	4.11	4.37
Asp186 OD2	2.10	3.24	1.64	2.09	2.64	2.28	4.92	2.84	4.67	2.79	2.75	2.48	3.21	3.55	3.13	3.64	3.32	2.62	3.81	4.31	3.48	3.74

The measured distances between the unliganded (1dlo) structure and each RT structure for all the atoms of the residues of polymerase active site.

Appendix 7.8 Displacement data for the primer grip

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2
Phe227 N	0.92	1.62	0.75	4.47	4.49	4.32	4.58	2.09	3.51	3.08	2.72
Phe227 CA	0.88	1.50	0.85	4.94	4.85	4.59	4.90	2.70	3.94	3.67	3.20
Phe227 C	0.91	1.45	0.70	4.88	4.80	4.47	4.80	2.73	3.83	3.67	3.10
Phe227 O	0.76	1.42	0.72	4.67	4.59	4.12	4.55	2.46	3.66	3.32	2.84
Phe227 CB	0.84	1.36	0.95	5.10	4.77	4.57	5.05	2.87	4.22	3.90	3.58
Phe227 CG	0.67	1.55	0.92	5.48	6.26	4.63	5.12	2.93	4.47	4.10	3.70
Phe227 CD1	0.87	1.78	1.00	6.09	6.01	4.87	5.28	3.12	4.76	4.45	3.70
Phe227 CD2	0.39	1.56	0.82	5.35	8.36	4.49	5.05	2.91	4.53	4.12	3.87
Phe227 CE1	0.75	1.96	0.97	6.51	7.60	4.99	5.36	3.19	5.01	4.70	3.80
Phe227 CE2	0.26	1.80	0.78	5.85	10.12	4.58	5.11	2.97	4.83	4.44	4.02
Phe227 CZ	0.42	1.97	0.85	6.39	9.67	4.84	5.26	3.07	5.01	4.67	3.94
Leu228 N	1.09	1.49	0.57	5.05	4.94	4.79	5.01	3.21	4.11	4.15	3.41
Leu228 CA	1.08	1.47	0.33	5.04	4.96	4.77	5.00	3.51	4.25	4.34	3.56
Leu228 C	1.05	1.31	0.27	5.21	5.14	5.01	5.40	4.10	4.78	4.96	4.06
Leu228 O	1.11	1.22	0.71	5.97	6.19	5.54	6.07	4.96	5.49	5.31	4.53
Leu228 CB	1.19	1.59	0.28	5.17	4.77	5.01	5.02	3.62	4.41	4.59	3.56
Leu228 CG	0.87	1.11	0.87	6.85	5.23	4.90	4.73	3.21	4.34	4.44	3.30
Leu228 CD1	0.75	3.03	1.89	7.54	5.45	5.92	5.34	3.94	5.67	5.69	2.08
Leu228 CD2	1.48	1.64	0.88	7.09	4.44	3.61	3.71	2.78	2.04	3.53	3.61
Trp229 N	0.99	1.21	0.81	4.59	4.22	4.71	5.12	3.67	4.44	5.14	3.89
Trp229 CA	0.97	1.13	0.81	4.65	4.21	4.88	5.36	4.11	4.85	5.71	4.32
Trp229 C	1.19	1.23	0.80	4.68	4.27	4.73	5.07	4.41	4.82	5.93	4.54
Trp229 O	1.45	1.35	0.82	4.65	4.44	4.36	4.75	4.23	4.59	5.72	4.38
Trp229 CB	0.85	1.10	0.77	4.58	4.29	4.77	5.32	4.10	4.80	5.71	4.33
Trp229 CG	0.99	1.07	0.75	4.77	4.52	4.98	5.44	4.19	5.21	5.25	4.50
Trp229 CD1	1.01	1.08	0.90	5.53	5.45	5.46	5.63	4.51	5.59	5.36	4.56
Trp229 CD2	1.20	1.14	0.74	5.14	4.77	4.72	5.45	4.05	5.26	5.06	4.60
Trp229 NE1	1.17	1.16	1.00	5.54	5.28	5.49	5.76	4.57	5.85	5.26	4.67
Trp229 CE2	1.26	1.20	0.88	4.98	4.64	5.06	5.66	4.29	5.67	5.08	4.70

lrt3	lrt4	lrt5	lrt6	lrt7	lrth	lrti	ltvr	luwb	lvrt	lvru
2.89	2.62	2.64	2.89	3.06	2.61	2.22	5.50	5.07	3.43	3.35
3.48	3.17	3.30	3.48	3.65	3.08	2.73	5.77	5.32	3.89	3.82
3.42	3.15	3.26	3.48	3.68	2.98	2.77	5.71	5.26	3.72	3.63
3.08	2.79	2.92	3.12	3.34	2.66	2.55	5.47	5.06	3.32	3.23
3.85	3.39	3.59	3.66	3.83	3.35	3.07	5.61	5.27	4.16	4.12
3.98	3.76	3.95	3.95	4.03	3.92	3.26	5.75	5.41	4.38	4.31
4.18	4.24	4.56	4.47	4.53	4.68	3.53	6.23	5.78	4.70	4.67
4.10	3.94	3.90	3.94	3.92	3.93	3.42	5.50	5.22	4.49	4.25
4.32	4.67	4.99	4.81	4.80	5.28	3.76	6.45	5.95	4.95	4.87
4.30	4.52	4.44	4.36	4.25	4.64	3.78	5.70	5.38	4.78	4.52
4.31	4.78	4.92	4.71	4.63	5.23	3.85	6.18	5.73	4.93	4.77
3.85	3.64	3.76	3.96	4.15	3.33	3.19	5.93	5.43	4.14	4.05
4.00	3.85	3.94	4.15	4.34	3.44	3.43	5.98	5.45	4.15	4.08
4.41	4.39	4.46	4.71	4.96	3.84	3.90	6.50	5.76	4.58	4.54
5.11	4.85	4.85	5.32	5.58	4.31	4.49	8.19	6.58	5.25	5.16
4.20	4.03	4.23	4.42	4.60	3.64	3.54	5.97	5.48	4.44	4.39
4.38	5.32	4.39	4.60	4.48	4.93	4.73	6.10	5.17	4.64	4.32
5.69	7.17	5.59	5.90	5.45	6.62	5.89	5.45	5.80	5.38	4.12
2.27	3.93	3.70	3.89	3.83	3.41	3.97	6.60	4.10	4.31	4.42
4.06	4.40	4.53	4.56	4.89	3.74	3.68	5.47	5.23	4.34	4.27
4.32	4.85	4.99	5.04	5.43	4.05	4.08	5.75	5.43	4.67	4.62
4.36	5.08	5.21	5.29	5.72	4.20	4.28	5.36	5.30	4.57	4.46
4.16	4.97	5.05	5.12	5.57	4.01	4.17	4.99	5.04	4.39	4.17
4.18	4.89	4.98	5.12	5.55	4.00	4.02	5.87	5.39	4.57	4.55
4.32	5.10	5.10	5.34	5.55	4.15	4.46	5.83	5.59	4.82	4.87
4.24	6.59	6.59	6.91	6.90	4.22	4.78	6.15	5.69	5.01	5.11
4.55	4.51	4.36	4.55	4.81	4.14	4.66	5.49	5.70	4.85	4.97
4.39	6.14	6.03	6.43	6.29	4.20	5.15	6.04	5.84	5.12	5.33
4.59	4.86	4.66	4.91	4.98	4.16	5.09	5.64	5.85	5.03	5.27

Appendix 7.8 Displacement data for the primer grip (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2
Trp229 CE3	1.38	1.24	0.73	6.58	6.19	4.24	5.31	3.77	5.00	5.15	4.64
Trp229 CZ2	1.46	1.33	0.93	5.26	4.98	4.94	5.74	4.29	5.82	5.22	4.85
Trp229 CZ3	1.58	1.36	0.82	7.48	7.27	4.11	5.38	3.75	5.15	5.26	4.77
Trp229 CH2	1.61	1.41	0.89	6.50	6.38	4.46	5.60	4.01	5.56	5.30	4.88
Met230 N	1.16	1.22	0.79	4.74	4.11	5.01	5.24	4.91	5.04	6.45	4.95
Met230 CA	1.31	1.26	0.87	4.81	4.21	4.98	5.12	5.32	5.12	6.86	5.34
Met230 C	1.25	1.42	1.27	4.41	3.93	4.52	4.64	4.97	4.71	6.47	4.69
Met230 O	1.26	1.51	1.44	4.03	3.53	3.99	4.02	4.53	4.27	5.99	3.77
Met230 CB	1.46	1.34	0.78	5.07	4.47	4.89	5.02	5.58	5.22	7.23	6.05
Met230 CG	1.65	1.47	0.66	4.78	4.62	4.65	5.25	6.04	6.05	7.75	5.79
Met230 SD	1.81	0.71	1.24	6.29	6.44	5.89	6.79	5.16	6.03	6.64	5.01
Met230 CE	2.46	3.43	1.90	8.11	8.26	7.29	7.81	4.29	7.79	5.89	5.05
Gly231 N	1.20	1.61	1.49	4.44	4.03	4.67	4.90	5.29	4.89	6.81	5.30
Gly231 CA	1.16	1.83	1.83	4.09	3.72	4.28	4.51	5.20	4.60	6.69	5.14
Gly231 C	1.25	2.20	1.92	3.41	3.19	3.39	3.68	4.45	3.78	5.62	4.23
Gly231 O	1.64	3.49	2.60	3.00	2.94	2.99	3.14	4.70	3.64	5.71	4.27
Tyr232 N	1.04	1.48	1.32	3.81	3.86	3.50	4.01	3.75	3.45	4.85	3.66
Tyr232 CA	1.17	1.53	1.27	3.34	3.52	2.87	3.42	3.07	2.80	3.89	2.96
Tyr232 C	1.15	1.61	1.27	3.57	3.75	3.04	3.68	2.65	3.00	3.41	2.68
Tyr232 O	1.26	1.82	1.81	3.57	3.79	3.04	4.14	2.60	3.74	3.74	2.99
Tyr232 CB	1.38	1.51	1.19	3.08	3.16	2.64	3.09	3.08	2.51	3.84	3.08
Tyr232 CG	1.43	1.41	1.25	2.82	2.75	2.45	2.86	3.65	2.32	4.41	2.81
Tyr232 CD1	1.48	1.81	1.51	3.92	3.77	1.99	3.71	3.84	2.06	3.14	1.54
Tyr232 CD2	1.43	3.40	1.50	3.25	3.10	2.84	3.50	4.13	2.51	6.68	4.84
Tyr232 CE1	1.53	1.81	1.83	3.76	3.47	1.91	3.51	4.49	2.02	3.58	1.46
Tyr232 CE2	1.43	3.36	1.40	3.31	3.10	2.76	3.53	4.73	2.46	7.36	4.63
Tyr232 CZ	1.49	1.27	1.52	2.65	2.28	2.30	2.62	4.90	2.22	5.65	2.34
Tyr232 OH	1.60	1.36	1.81	2.67	2.16	2.30	2.55	5.58	2.23	6.32	2.07
Glu233 N	1.11	1.48	1.24	3.85	3.90	3.28	3.42	2.46	2.38	2.74	2.22

lrt3	lrt4	lrt5	lrt6	lrt7	lrth	lrti	ltvr	luwb	lvrt	lvru
4.74	5.09	4.87	4.89	5.28	4.13	4.52	5.16	5.68	4.75	4.86
4.81	4.89	4.61	4.66	4.83	4.15	5.39	5.45	5.99	5.12	5.44
4.96	6.24	6.00	5.82	6.26	4.12	4.81	4.97	5.82	4.85	5.04
5.00	5.81	5.55	5.30	5.73	4.12	5.24	5.11	5.97	5.03	5.33
4.63	5.41	5.61	5.76	6.19	4.63	4.60	5.49	5.42	4.73	4.73
4.73	5.72	5.91	6.16	6.64	4.94	4.87	5.34	5.38	4.71	4.74
4.36	5.42	5.64	5.84	6.25	4.43	4.42	5.01	4.89	4.44	4.41
3.81	5.08	5.36	5.37	5.78	3.77	3.88	5.04	4.39	4.22	3.93
4.69	5.86	5.97	6.40	7.03	5.24	5.07	5.18	5.31	4.60	4.60
4.67	6.51	6.39	6.98	7.70	5.26	5.27	5.24	5.16	4.48	6.07
6.07	5.71	5.57	5.99	6.52	5.11	5.22	6.46	6.30	5.99	6.07
7.73	5.47	5.49	4.97	5.66	4.30	7.52	8.29	7.84	7.14	8.24
4.75	5.62	5.84	6.27	6.55	4.82	4.72	4.72	4.91	4.45	4.67
4.55	5.55	5.74	6.19	6.45	4.68	4.49	4.38	4.47	4.28	4.42
3.79	4.66	4.78	5.23	5.42	3.92	3.80	3.86	3.60	3.58	3.67
3.78	4.75	4.91	5.38	5.60	4.13	4.02	3.49	3.01	3.42	3.57
3.46	4.23	4.19	4.55	4.61	3.40	3.32	3.94	4.05	3.32	3.28
2.82	3.45	3.33	3.70	3.67	2.78	2.71	3.65	3.43	2.77	2.80
2.95	3.10	2.99	3.19	3.20	2.57	2.70	3.99	3.75	2.99	3.00
3.66	3.39	3.25	3.31	3.35	2.80	3.05	4.63	4.19	3.46	3.49
2.50	3.42	3.27	3.73	3.60	2.76	2.51	3.36	3.04	2.45	2.50
2.38	4.00	3.91	4.37	4.25	3.13	2.65	2.92	2.72	2.25	2.35
2.29	2.54	2.50	2.93	2.95	1.98	2.57	2.51	3.65	2.03	2.14
2.43	6.37	6.31	6.71	6.55	5.24	2.99	3.13	3.30	2.44	2.57
2.28	2.98	3.01	3.38	3.44	2.12	2.85	2.37	3.45	2.05	2.18
2.42	7.02	7.04	7.39	7.26	5.70	3.23	3.01	3.22	2.44	2.59
2.36	5.20	5.27	5.60	5.56	3.90	3.16	2.66	2.29	2.26	2.41
2.42	5.85	5.99	6.25	6.24	4.35	3.49	2.71	2.13	2.37	2.49
2.31	2.62	2.51	2.75	2.79	2.25	2.34	3.71	3.54	2.65	2.54

Appendix 7.8 Displacement data for the primer grip (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2	1rt3	1rt4	1rt5	1rt6	1rt7	1rth	1rti	1tvr	1uwb	1vrt	1vru
Glu233 CA	1.20	1.61	1.17	4.05	4.07	3.39	3.61	2.13	2.47	2.33	1.99	2.36	2.40	2.30	2.35	2.43	2.20	2.33	3.92	3.75	2.76	2.61
Glu233 C	1.13	1.61	1.22	3.77	3.83	3.11	3.37	1.82	2.14	1.98	1.69	1.97	1.91	1.83	1.93	2.04	1.83	2.05	3.73	3.53	2.48	2.30
Glu233 O	1.15	1.86	1.45	3.35	3.57	2.69	2.98	1.84	1.59	1.80	1.39	1.45	1.74	1.72	1.79	1.91	1.65	1.97	3.41	3.17	2.07	1.83
Glu233 CB	1.24	1.74	1.17	4.19	4.24	3.42	3.67	2.01	2.42	1.95	1.67	2.32	2.27	2.19	2.06	2.21	2.11	2.31	4.11	3.86	2.68	2.52
Glu233 CG	1.46	3.38	0.59			2.79	2.62	2.95	2.01	3.10	2.84	2.14	2.11	2.09	2.93	2.92	2.08	2.07	4.27	2.93	2.06	2.06
Glu233 CD	1.80	4.20	1.49			2.50	1.94	2.87	2.26	2.79	2.32	2.71	2.59	2.65	2.56	2.76	2.63	2.59	3.74	2.17	2.07	2.18
Glu233 OE1	2.03	3.92	2.63			2.65	2.25	2.53	2.13	2.76	1.40	1.78	1.70	1.75	2.23	2.67	2.12	2.21	3.07	1.53	1.80	1.91
Glu233 OE2	2.32	5.20	2.02			2.53	3.18	3.67	2.92	3.15	3.99	4.82	4.63	4.74	3.21	3.16	4.53	3.87	4.04	2.37	3.17	3.23
Leu234 N	0.96	1.32	1.06	4.03	3.94	3.40	3.63	1.69	2.73	2.17	1.99	2.39	2.12	2.06	2.04	2.05	2.08	2.14	3.99	3.77	2.83	2.63
Leu234 CA	0.90	1.36	1.08	3.87	3.76	3.32	3.48	1.47	2.70	1.93	1.89	2.33	1.84	1.80	1.78	1.78	1.98	1.94	3.86	3.64	2.72	2.59
Leu234 C	0.80	1.38	1.07	3.97	3.81	3.34	3.63	1.11	2.70	1.76	1.85	2.30	1.67	1.65	1.52	1.45	1.90	1.83	3.98	3.75	2.71	2.57
Leu234 O	0.81	1.11	1.07	4.73	4.68	4.15	4.44	1.38	3.26	2.48	2.32	2.81	2.36	2.38	2.32	2.12	2.54	2.39	4.70	4.53	3.33	3.11
Leu234 CB	0.87	1.30	1.08	4.13	3.96	3.71	3.75	1.74	3.19	2.30	2.23	2.88	2.39	2.34	2.29	2.33	2.60	2.17	4.13	3.94	3.13	3.13
Leu234 CG	0.84	1.67	1.18	3.71	4.03	3.49	3.69	2.04	2.81	2.53	2.17	2.60	2.13	2.01	2.60	2.59	2.10	2.03	4.09	4.04	2.78	3.24
Leu234 CD1	0.97	1.82	0.86	4.40	5.11	3.85	4.84	3.98	3.97	5.06	4.34	4.90	4.75	4.66	5.14	4.85	4.59	3.90	5.00	5.50	4.55	3.60
Leu234 CD2	0.72	2.15	1.37	2.64	3.23	2.83	2.72	0.53	4.44	3.10	3.41	3.00	2.36	2.30	3.11	3.49	2.85	3.68	3.34	3.37	4.33	3.06
His235 N	1.01	1.92	1.24	3.20	2.91	2.74	3.10	0.54	2.30	0.95	1.37	1.82	1.05	1.10	0.97	0.69	1.43	1.05	3.33	3.19	2.15	2.03
His235 CA	1.16	2.07	1.27	3.25	2.99	2.98	3.33	0.11	2.50	1.10	1.45	2.11	1.47	1.58	1.64	1.36	2.01	0.95	3.38	3.43	2.45	2.32
His235 C	0.92	2.11	0.92	3.44	3.16	3.19	3.62	0.39	2.95	1.44	1.73	2.71	1.92	2.07	2.11	1.77	2.53	1.14	3.48	3.69	3.06	2.81
His235 O	0.63	2.19	0.80	3.31	2.96	2.94	3.62	0.56	3.01	1.23	1.90	2.60	1.57	1.90	1.49	1.14	2.24	1.24	3.35	3.35	3.19	2.75
His235 CB	1.58	2.30	1.43	2.82	2.58	2.67	3.00	0.44	2.13	0.79	1.08	1.79	1.42	1.56	1.77	1.66	2.05	0.61	3.00	3.10	2.08	1.99
His235 CG	1.74	2.42	1.38	2.65	2.42	2.46	2.72	0.69	1.58	0.22	0.78	1.36	1.11	1.22	1.30	1.31	1.62	0.48	2.82	2.78	1.45	1.50
His235 ND1	1.63	2.50	1.34	1.98	1.84	2.46	2.38	2.40	2.65	2.11	1.46	3.50	2.20	1.81	3.01	3.38	2.26	1.00	2.12	2.76	2.94	3.39
His235 CD2	2.08	2.54	1.35	3.11	2.88	2.18	2.70	1.31	2.02	2.23	0.72	0.92	1.06	1.04	1.78	1.20	1.23	1.28	3.22	2.51	1.69	1.21
His235 CE1	1.86	2.71	1.29	2.07	2.06	2.19	2.15	2.23	2.15	1.19	0.93	3.22	2.05	1.86	2.16	2.67	2.18	0.54	2.12	2.44	2.31	2.83
His235 NE2	2.17	2.75	1.30	2.74	2.67	2.02	2.35	1.17	0.73	1.62	0.68	0.86	0.56	0.92	1.22	0.86	1.07	1.18	2.79	2.31	0.39	0.33

The measured distances between the unliganded (1dlo) structure and each RT structure for all the atoms of the residues of primer grip.

Appendix 7.9 Distance between the primer grip and the PAS

PDB file	Asp110	Asp185	Asp186
1dlo	13.11	12.19	10.07
1rtd	13.59	14.59	11.62
2hmi	14.02	13.96	11.34
1hmv	12.41	12.29	10.34
1bqm	17.74	15.50	14.42
1bqn	16.71	14.65	13.32
1hni	18.25	16.59	14.97
1hnv	18.74	17.27	15.78
1klm	19.71	18.55	17.10
1rev	19.68	18.24	16.85
1rt1	22.13	20.04	18.70
1rt2	20.57	18.08	17.12
1rt3	18.76	17.06	15.83
1rt4	20.58	19.62	17.67
1rt5	21.01	20.02	18.12
1rt6	20.54	19.88	18.18
1rt7	21.35	20.14	18.29
1rth	20.31	18.50	16.98
1rti	18.90	17.81	15.89
1tvr	18.09	16.78	15.54
1uwb	18.28	17.50	15.86
1vrt	19.00	17.43	16.17
1vru	19.65	17.83	16.34

The measured distance between C_α of Gly231 of the primer grip and C_α of the residues of the catalytic triad in each of the crystal structures.

Appendix 7.10 Displacement of the thumb

PDB file	286:C_α	296:C_α	306:C_α
1rtd	18.37	19.13	12.99
2hmi	21.40	19.13	14.11
1hmv	3.10	3.43	2.63
1bqm	18.77	19.30	13.18
1bqn	19.19	19.49	13.37
1hni	19.94	21.46	13.60
1hnv	19.60	20.96	13.54
1klm	24.10	23.79	15.37
1rev	25.40	25.33	15.70
1rt1	24.40	24.10	15.69
1rt2	24.00	22.86	14.26
1rt3	24.91	25.23	15.90
1rt4	24.12	23.81	15.75
1rt5	24.15	23.86	15.80
1rt6	24.33	24.31	16.16
1rt7	24.54	24.55	16.24
1rth	24.36	23.86	15.40
1rti	23.90	23.45	15.05
1tvr	19.38	19.78	13.24
1uwb	20.29	20.70	13.64
1vrt	25.73	25.59	15.94
1vru	25.53	25.15	16.12

The measured distances between the unliganded (1dlo) structure and each RT structure at three residues (C_α atoms of 286, 296 and 306) representing three different parts of the thumb (Figure 3.29).

Appendix 7.11 Displacement data for the template grip

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2
Lys73 N	4.15	1.37		4.80	4.73	3.07	5.03	3.23	6.81	3.26	3.94
Lys73 CA	3.82	2.24		4.52	4.57	3.48	5.21	3.27	7.04	3.28	4.05
Lys73 C	3.52	3.06		4.48	4.58	3.28	4.81	3.23	6.72	3.22	3.77
Lys73 O	3.60	4.03		3.80	3.82	2.72	3.74	2.81	6.08	2.94	3.23
Lys73 CB	3.68	2.80		4.48	4.57	3.48	5.07	3.27	6.99	3.23	4.06
Lys73 CG	2.75			4.74	4.78	3.86	5.08	4.19	7.22	4.18	4.07
Lys73 CD	2.06			4.46	4.32	3.83	4.75	3.34	7.27	3.26	4.06
Lys73 CE	1.24			5.18	4.92	4.87	5.94	3.98	7.82	3.40	4.35
Lys73 NZ	1.98			5.08	4.92	4.04	5.17	2.25	5.96	1.82	2.71
Leu74 N	3.22	3.00		5.39	5.61	3.78	5.84	3.81	7.18	3.71	4.31
Leu74 CA	2.98	3.83		5.52	5.85	3.67	5.67	3.82	6.97	3.70	4.17
Leu74 C	2.54	3.28		5.37	5.64	3.84	5.92	3.86	6.99	3.82	4.12
Leu74 O	2.79	3.25		5.12	5.73	4.01	6.88	4.12	7.44	4.13	4.18
Leu74 CB	3.17	4.65		5.54	5.92	4.05	6.33	3.83	7.09	3.72	4.24
Leu74 CG	3.56			6.20	6.38	3.80	6.18	3.80	6.84	3.75	4.71
Leu74 CD1	2.02			5.41	7.69	3.68	5.91	4.54	7.10	3.96	5.81
Leu74 CD2	6.15			7.60	6.44	3.92	6.63	3.86	6.80	3.67	4.81
Val75 N	2.32	2.84	0.71	5.54	5.27	3.83	5.09	3.58	6.52	3.61	4.06
Val75 CA	1.87	2.27	0.63	5.46	5.01	3.98	5.21	3.55	6.45	3.63	3.95
Val75 C	1.80	2.20	0.71	5.60	5.01	4.52	5.44	3.57	6.34	3.62	3.90
Val75 O	2.28	2.41	1.08	6.31	5.49	5.79	5.76	3.79	6.28	3.68	3.96
Val75 CB	1.71	2.16	0.51	5.53	5.08	4.14	5.33	3.56	6.27	3.58	3.82
Val75 CG1	1.18	1.41	0.51	5.40	4.75	4.16	5.19	3.40	6.18	3.42	3.67
Val75 CG2	1.93	2.62	0.64	5.54	5.31	3.99	5.59	3.62	6.40	3.63	3.90
Asp76 N	1.26	1.97	0.69	5.02	4.47	3.64	5.17	3.38	6.31	3.54	3.79
Asp76 CA	1.06	1.96	0.67	5.08	4.49	4.16	5.51	3.36	6.23	3.52	3.74
Asp76 C	0.91	1.76	0.51	4.99	4.52	3.75	5.41	3.25	5.89	3.28	3.55
Asp76 O	0.75	1.36	0.18	4.99	4.45	3.04	4.97	3.10	5.71	3.05	3.62
Asp76 CB	0.78	2.06	0.67	4.77	4.17	4.20	5.14	3.50	6.66	3.76	3.93

lrt3	lrt4	lrt5	lrt6	lrt7	lrth	lrti	ltvr	luwb	lvrt	lvru
6.13	4.53	4.75	3.18	3.13	4.26	4.00	5.97	4.90	6.68	6.72
6.37	4.39	4.68	3.26	3.15	4.09	4.06	6.31	5.23	6.87	6.89
6.07	4.26	4.52	3.09	3.10	3.95	3.96	5.95	5.00	6.60	6.54
5.28	3.86	4.01	2.57	2.75	3.60	3.53	4.78	4.03	6.07	5.74
6.38	4.27	4.58	3.26	3.15	3.91	4.04	6.13	5.23	6.85	6.85
6.60	5.21	5.43	3.34	4.19	5.04	4.89	6.64	5.45	6.94	6.89
6.66	4.52	4.69	3.36	2.97	4.22	4.13	6.54	4.66	6.98	6.73
7.01	4.56	4.84	3.37	3.09	4.32	3.98	7.28	5.65	7.37	6.99
5.13	2.42	2.69	1.15	1.66	2.55	3.49	5.79	4.80	5.74	6.60
6.74	4.80	5.09	3.76	3.70	4.48	4.58	7.09	6.01	7.01	7.19
6.57	4.73	5.01	3.70	3.73	4.41	4.57	6.89	5.93	6.86	6.98
6.68	4.83	5.07	3.84	3.86	4.45	4.56	6.97	6.10	7.04	7.20
7.42	5.22	5.46	4.41	4.31	4.68	4.81	7.56	6.72	7.88	8.14
6.67	4.79	5.08	3.74	3.74	4.49	4.58	7.22	6.08	7.01	7.23
6.47	5.22	5.14	3.66	3.72	4.71	4.64	7.14	6.19	7.05	8.06
6.75	6.57	6.04	4.02	4.17	5.60	5.25	7.06	6.33	8.31	9.00
6.41	5.09	5.22	3.60	3.64	4.93	4.65	7.23	6.98	6.86	8.21
5.93	4.45	4.72	3.32	3.45	4.22	4.27	6.35	5.49	6.30	6.38
5.91	4.45	4.70	3.38	3.47	4.18	4.20	6.32	5.51	6.34	6.44
5.86	4.46	4.69	3.48	3.52	4.17	4.17	6.73	5.94	6.28	6.39
5.94	4.55	4.78	3.74	3.75	4.19	4.21	8.01	6.76	6.27	6.29
5.78	4.43	4.66	3.46	3.52	4.12	4.13	6.50	5.73	6.20	6.25
5.84	4.36	4.58	3.49	3.55	3.92	4.08	6.28	5.56	6.21	6.42
5.77	4.46	4.69	3.40	3.45	4.19	4.15	6.46	5.71	6.22	6.16
5.71	4.39	4.59	3.32	3.35	4.15	4.10	5.71	5.31	6.19	6.44
5.66	4.39	4.55	3.38	3.34	4.12	4.06	6.15	5.79	6.14	6.43
5.53	4.23	4.36	3.29	3.17	3.88	3.89	5.86	5.63	5.95	6.11
5.49	4.08	4.17	3.29	2.98	3.69	3.73	5.42	5.21	6.17	6.15
5.91	4.56	4.75	3.46	3.50	4.38	4.29	5.92	5.51	6.49	6.96

Appendix 7.11 Displacement data for the template grip (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2
Asp76 CG	0.53	2.25	0.62	4.61	3.94	5.06	5.86	3.30	6.45	3.86	3.93
Asp76 OD1	0.65	2.25	0.82	5.11	4.27	5.58	6.79	3.33	6.22	3.81	3.69
Asp76 OD2	0.31	2.55	0.56	4.27	3.86	5.33	5.80	3.17	6.60	4.13	4.30
Phe77 N	1.15	2.16	0.77	4.91	4.73	4.26	5.88	3.29	5.75	3.22	3.22
Phe77 CA	1.20	2.13	0.71	4.73	4.72	4.15	5.91	3.15	5.42	3.00	3.00
Phe77 C	1.26	2.16	0.70	4.44	4.54	4.15	5.95	3.15	5.38	3.00	3.01
Phe77 O	1.31	2.20	0.81	3.50	3.82	3.99	6.07	3.11	5.25	3.04	2.97
Phe77 CB	1.38	2.09	0.71	4.78	4.79	4.85	6.06	2.93	5.19	2.87	2.78
Phe77 CG	1.53	2.12	0.63	4.95	4.86	4.43	5.74	2.98	5.38	2.90	2.81
Phe77 CD1	1.44	2.04	0.51	5.02	4.85	3.18	4.33	3.03	5.31	2.85	2.89
Phe77 CD2	1.90	2.28	0.74	5.06	4.93	6.10	7.40	2.98	5.64	2.99	2.80
Phe77 CE1	1.61	2.10	0.54	5.16	4.88	3.04	4.19	3.07	5.52	2.88	2.93
Phe77 CE2	2.07	2.29	0.74	5.20	4.96	5.70	7.06	3.02	5.83	3.02	2.83
Phe77 CZ	1.88	2.19	0.67	5.23	4.93	3.75	5.18	3.06	5.78	2.97	2.90
Arg78 N	1.27	2.15	0.62	5.35	5.14	4.39	5.86	3.20	5.51	2.99	3.08
Arg78 CA	1.27	2.18	0.66	5.14	5.03	4.33	5.84	3.15	5.45	2.96	3.09
Arg78 C	1.27	2.16	0.71	5.00	4.97	4.24	5.76	3.03	5.42	2.91	3.11
Arg78 O	1.42	2.27	0.82	4.81	5.13	4.34	5.67	3.03	5.53	2.88	3.14
Arg78 CB	1.15	2.40	0.80	5.52	5.70	4.25	5.79	3.57	5.77	3.15	3.28
Arg78 CG	1.20		1.33	7.70	8.18	4.30	5.77	3.97	6.34	3.36	3.74
Arg78 CD	1.06		1.72	9.58	8.76	6.35	7.37	3.86	6.67	3.52	3.87
Arg78 NE	0.90		0.83	9.61	8.96	6.67	7.80	3.77	5.97	3.66	3.90
Arg78 CZ	0.70		1.08	9.63	9.33	9.03	10.12	3.71	6.72	3.82	3.95
Arg78 NH1	0.64		1.93	9.62	9.58	10.78	11.67	3.79	8.18	3.86	3.97
Arg78 NH2	0.61		2.35	9.68	9.46	10.07	11.28	3.71	6.19	3.95	4.00
Glu79 N	1.08	2.07	0.57	5.12	4.90	3.98	5.75	2.99	5.27	2.87	3.12
Glu79 CA	1.09	2.07	0.66	5.10	4.88	3.94	5.64	2.93	5.25	2.89	3.16
Glu79 C	1.21	1.90	0.64	4.91	4.65	3.84	5.50	2.77	5.01	2.70	2.95
Glu79 O	1.27	1.77	0.73	4.94	4.59	3.69	5.40	2.53	4.67	2.45	2.70

lrt3	lrt4	lrt5	lrt6	lrt7	lrth	lrti	ltvr	luwb	lvrt	lvru
5.72	4.53	4.65	3.42	3.47	4.36	4.21	6.91	6.33	6.61	6.75
5.46	4.39	4.57	3.34	3.39	4.27	4.16	7.71	7.29	6.37	6.50
5.85	4.68	4.65	3.80	3.78	4.47	4.24	7.29	6.31	6.95	6.84
5.46	4.18	4.34	3.16	3.15	3.80	3.87	6.27	6.06	5.53	5.71
5.31	4.01	4.13	3.02	2.98	3.57	3.70	6.17	6.11	5.32	5.38
5.30	4.04	4.10	3.02	3.04	3.55	3.72	6.03	6.12	5.36	5.35
5.19	4.06	4.02	3.03	3.17	3.46	3.69	5.91	6.23	5.26	5.27
5.08	3.77	3.96	2.84	2.73	3.41	3.48	6.70	6.68	5.12	5.01
5.22	3.78	3.99	2.89	2.74	3.42	3.50	6.51	6.04	5.20	5.02
5.20	3.81	4.00	2.85	2.77	3.39	3.55	5.06	4.39	5.31	5.16
5.39	3.77	4.01	2.99	2.73	3.48	3.46	8.23	7.56	5.21	4.90
5.33	3.81	4.02	2.89	2.77	3.39	3.55	4.91	3.98	5.45	5.18
5.52	3.80	4.04	3.05	2.74	3.51	3.47	8.02	6.91	5.30	4.91
5.49	3.82	4.05	3.00	2.77	3.46	3.52	6.13	4.85	5.42	5.06
5.43	4.05	4.17	3.00	2.99	3.65	3.81	6.03	6.00	5.53	5.41
5.37	4.05	4.13	2.94	2.99	3.63	3.82	5.98	5.94	5.59	5.36
5.33	3.95	4.02	2.90	2.97	3.54	3.85	5.67	5.88	5.64	5.37
5.36	4.00	4.11	2.90	3.11	3.66	4.05	5.31	6.02	5.61	5.30
5.63	4.30	4.38	3.19	3.19	3.89	4.12	5.83	5.90	5.95	5.55
6.68	4.40	4.67	3.44	3.35	4.13	4.53	6.47	5.88	6.35	7.62
6.47	4.33	4.23	3.58	3.45	3.98	4.90	8.42	7.46	6.77	8.13
6.04	4.27	3.67	3.69	3.54	3.86	3.89	9.71	7.99	6.06	6.35
5.84	4.98	4.76	3.81	3.63	4.69	4.73	12.07	10.31	6.93	6.75
6.06	6.11	6.72	3.83	3.65	5.96	6.46	13.10	11.82	8.52	9.08
5.40	4.81	4.09	3.89	3.71	4.44	4.34	13.62	11.48	6.44	5.14
5.23	3.81	3.84	2.87	2.83	3.32	3.66	5.80	5.71	5.72	5.46
5.19	3.75	3.79	2.91	2.87	3.29	3.73	5.64	5.69	5.74	5.49
5.05	3.61	3.66	2.85	2.79	3.23	3.56	5.41	5.57	5.53	5.30
4.91	3.39	3.45	2.66	2.63	3.12	3.44	4.85	5.40	5.36	5.23

Appendix 7.11 Displacement data for the template grip (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2
Glu79 CB	0.74	2.20	0.77	5.35	5.10	4.04	5.81	3.08	5.50	3.13	3.35
Glu79 CG	1.71	2.77	1.25	5.62	4.67	5.27	6.57	3.50	5.86	3.87	3.51
Glu79 CD	2.68	2.97	0.88	6.32	6.81	6.19	7.01	3.34	6.01	3.53	4.29
Glu79 OE1	2.94	3.54	2.93	5.99	8.84	6.18	6.90	3.85	6.04	3.55	6.07
Glu79 OE2	3.54	3.07	1.50	7.47	6.98	7.09	7.50	4.06	6.12	4.50	3.79
Leu80 N	1.24	1.90	0.83	4.75	4.52	3.89	5.52	2.93	5.20	2.85	3.05
Leu80 CA	1.38	1.72	0.77	4.55	4.28	3.70	5.35	2.77	4.99	2.67	2.83
Leu80 C	1.58	1.65	0.72	4.27	3.99	3.49	5.09	2.61	4.76	2.55	2.61
Leu80 O	1.84	1.77	0.69	4.03	3.76	3.30	4.80	2.38	4.53	2.31	2.32
Leu80 CB	1.33	1.43	0.67	4.45	4.21	3.72	5.30	2.73	5.04	2.65	2.74
Leu80 CG	1.36	1.87	0.43	4.35	4.01	3.48	4.98	2.48	4.87	2.40	2.58
Leu80 CD1	1.39	2.89	0.59	4.44	4.11	3.56	5.15	2.58	4.90	2.52	2.54
Leu80 CD2	1.25	2.31	0.82	4.29	4.07	3.52	4.86	2.49	4.91	2.37	2.71
Asn81 N	1.49	1.68	0.77	4.32	4.03	3.50	5.18	2.76	4.84	2.74	2.81
Asn81 CA	1.64	1.62	0.76	4.11	3.82	3.32	4.96	2.66	4.67	2.69	2.67
Asn81 C	1.83	1.87	0.76	4.23	3.87	3.34	4.94	2.68	4.58	2.72	2.69
Asn81 O	2.27	2.29	1.07	4.19	3.75	3.19	4.90	3.14	4.71	3.17	2.97
Asn81 CB	1.75	1.51	0.73	4.09	3.81	3.30	4.97	2.64	4.76	2.71	2.72
Asn81 CG	1.60	1.04	0.70	3.59	3.39	3.03	4.86	2.54	4.75	2.61	2.58
Asn81 OD1	1.64	1.22	1.20	2.47	3.02	3.27	4.69	2.61	4.69	2.58	2.66
Asn81 ND2	2.01	2.86	1.71	5.33	3.81	3.08	5.34	2.66	5.12	2.86	2.53
Lys82 N	1.64	1.79	0.59	4.43	4.01	3.52	4.99	2.30	4.39	2.44	2.54
Lys82 CA	1.80	1.99	0.60	4.52	4.05	3.51	4.96	2.25	4.28	2.42	2.51
Lys82 C	1.86	2.02	0.38	4.53	3.91	3.47	4.86	2.20	4.11	2.36	2.44
Lys82 O	1.83	2.67	0.42	4.56	3.82	3.35	4.71	2.21	3.92	2.55	3.06
Lys82 CB	1.72	2.08	0.70	4.73	4.55	3.52	4.95	1.95	4.40	2.22	2.25
Lys82 CG	1.75	1.57	1.51	5.03	4.78	3.51	4.84	3.85	4.43	3.73	4.00
Lys82 CD	1.63	1.55	1.45	5.44	5.29	3.43	4.91	3.79	4.86	3.55	3.67
Lys82 CE	1.61	3.43	1.64	5.88	5.78	3.30	5.16	5.70	4.90	5.00	4.84

lrt3	lrt4	lrt5	lrt6	lrt7	lrth	lrti	ltvr	luwb	lvrt	lvru
5.35	3.86	3.92	3.06	3.02	3.34	3.97	5.98	5.87	5.95	5.70
6.13	4.16	4.19	3.18	3.49	3.55	4.27	7.24	6.74	6.45	6.17
6.51	4.16	4.25	3.37	3.27	3.56	4.34	8.06	7.26	6.66	6.31
6.36	5.04	4.13	4.81	3.76	3.43	4.34	8.74	7.38	6.87	6.25
7.12	4.38	4.42	3.23	3.95	3.71	4.40	8.27	7.64	6.69	6.53
5.11	3.74	3.78	3.05	2.93	3.33	3.55	5.83	5.66	5.57	5.28
4.94	3.60	3.62	2.96	2.82	3.18	3.35	5.66	5.51	5.33	5.04
4.69	3.49	3.49	2.88	2.72	3.03	3.16	5.48	5.31	4.97	4.68
4.41	3.29	3.22	2.72	2.53	2.72	2.87	5.27	5.05	4.62	4.22
4.94	3.54	3.59	2.84	2.75	3.13	3.30	5.85	5.47	5.15	5.07
4.71	3.22	3.25	2.53	2.48	2.74	3.03	5.47	4.67	5.04	4.95
4.67	3.11	3.22	2.67	2.57	2.54	2.96	5.31	4.90	4.91	4.86
4.86	3.30	3.31	2.52	2.48	2.86	3.14	5.82	4.58	5.28	5.17
4.79	3.66	3.72	3.03	2.90	3.26	3.36	5.61	5.50	5.08	4.92
4.60	3.60	3.64	2.98	2.85	3.21	3.24	5.51	5.36	4.79	4.65
4.64	3.61	3.62	2.99	2.89	3.18	3.21	5.56	5.37	4.70	4.67
4.88	3.84	3.73	3.48	3.39	3.28	3.17	6.06	5.62	4.73	4.91
4.57	3.62	3.68	2.97	2.81	3.30	3.32	5.47	5.31	4.87	4.69
4.37	3.61	3.63	2.96	2.76	3.25	3.33	5.33	5.23	4.85	4.62
4.35	3.60	3.66	3.03	2.80	3.23	3.34	5.43	5.28	4.88	4.50
4.61	3.84	3.75	3.17	2.90	3.44	3.59	5.39	5.38	5.08	5.09
4.47	3.45	3.50	2.54	2.50	3.08	3.24	5.12	5.09	4.58	4.56
4.47	3.42	3.45	2.47	2.46	3.03	3.21	5.14	5.07	4.48	4.58
4.33	3.25	3.27	2.39	2.41	2.82	2.97	5.13	5.08	4.24	4.37
4.21	3.13	3.21	2.37	2.40	2.82	2.81	5.11	5.23	4.01	4.18
4.69	3.47	3.50	2.12	2.16	3.11	3.34	4.72	4.95	4.72	4.92
4.47	3.53	3.38	3.83	3.83	3.29	3.31	5.68	4.85	4.86	5.00
4.80	3.79	4.04	3.62	3.62	3.49	3.34	5.17	5.14	5.39	5.27
4.86	3.92	4.62	3.29	5.18	3.72	3.28	6.73	5.19	6.27	5.39

Appendix 7.11 Displacement data for the template grip (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2
Lys82 NZ	2.40	2.92	2.40	6.01	5.96	3.51	4.72	5.99	4.84	5.29	4.50
Arg83 N	1.95	1.33	0.32	4.58	4.04	3.59	4.98	2.19	4.19	2.17	1.96
Arg83 CA	2.01	1.14	0.22	4.54	3.96	3.52	4.86	2.03	4.03	1.96	1.66
Arg83 C	2.05	1.00	0.16	4.33	3.64	3.35	4.65	1.98	3.78	1.92	1.82
Arg83 O	2.12	0.69	0.41	4.08	3.08	3.27	4.71	1.98	3.67	1.96	2.18
Arg83 CB	1.95	1.19	0.42	4.79	4.32	3.81	5.19	2.09	4.31	1.97	1.58
Arg83 CG	1.94	1.48	0.10	5.17	4.78	3.71	5.71	2.16	4.52	2.12	3.14
Arg83 CD	1.52	1.81	0.65	4.72	4.62	4.49	4.77	2.54	5.20	2.33	3.30
Arg83 NE	2.98	2.35	0.95	5.82	6.24	6.14	6.07	3.37	6.35	2.97	3.74
Arg83 CZ	2.67	1.65	1.06	4.69	5.60	6.13	6.21	2.63	5.34	2.67	3.19
Arg83 NH1	1.28	0.45	0.90	2.71	3.79	5.20	5.80	2.96	3.53	3.65	3.84
Arg83 NH2	4.22	2.20	1.35	5.85	7.20	7.59	7.40	3.30	6.50	2.99	3.54
Asp86 N	1.84	1.05	0.69	3.49	3.34	2.52	3.39	1.54	2.55	1.58	1.46
Asp86 CA	1.76	1.12	0.90	3.38	3.24	2.55	3.35	1.60	2.71	1.57	1.57
Asp86 C	1.70	0.88	0.22	3.11	2.98	2.42	3.22	1.40	2.52	1.29	1.32
Asp86 O	1.63	0.89	1.10	2.96	2.92	2.34	2.94	1.18	2.25	0.95	0.99
Asp86 CB	1.58	1.41	2.10	3.10	2.94	2.21	2.98	1.42	2.46	1.57	1.56
Asp86 CG	1.85	0.90	0.70	3.67	3.21	2.81	3.76	1.84	3.14	1.97	2.34
Asp86 OD1	2.69	1.62	2.20	3.31	3.12	2.60	4.38	1.39	2.72	2.07	3.10
Asp86 OD2	1.39	3.03	1.50	4.55	3.58	3.63	3.85	3.15	4.25	3.01	3.13
Phe87 N	1.72	0.72	0.73	3.06	2.83	2.43	3.58	1.52	2.67	1.42	1.59
Phe87 CA	1.62	0.77	0.45	2.79	2.55	2.31	3.49	1.38	2.52	1.23	1.49
Phe87 C	1.68	0.91	0.40	2.59	2.43	2.27	3.42	1.37	2.39	1.18	1.43
Phe87 O	2.20	1.49	0.47	2.62	2.45	2.28	3.42	1.36	2.44	1.22	1.54
Phe87 CB	1.56	0.60	0.78	2.95	2.70	2.66	3.85	1.66	2.87	1.42	1.97
Phe87 CG	1.51	0.49	0.64	3.13	2.80	2.52	3.74	1.58	2.83	1.39	1.86
Phe87 CD1	1.67	2.16	0.68	5.03	4.72	4.38	5.51	3.30	2.66	3.01	1.49

lrt3	lrt4	lrt5	lrt6	lrt7	lrth	lrti	ltvr	luwb	lvrt	lvru
3.70	3.11	3.84	2.31	5.43	2.90	3.56	6.56	4.97	6.33	5.00
4.39	3.23	3.23	2.38	2.39	2.71	2.93	5.14	5.04	4.33	4.42
4.21	2.99	2.99	2.20	2.23	2.47	2.61	5.10	4.99	4.10	4.19
3.89	2.79	2.85	2.16	2.15	2.33	2.35	5.07	4.69	3.83	3.81
3.75	2.72	2.92	2.11	2.04	2.43	2.04	5.28	4.61	3.74	3.60
4.55	3.13	3.13	2.26	2.27	2.54	2.79	5.15	5.29	4.47	4.59
4.96	3.36	3.34	2.31	2.36	2.78	3.06	5.19	6.04	4.81	5.80
5.69	3.76	3.82	2.66	2.56	3.24	3.54	5.77	4.85	5.71	4.79
6.95	4.55	4.61	3.74	3.25	4.05	4.38	6.88	6.19	7.01	6.43
6.04	3.87	3.88	2.86	2.68	3.58	3.76	7.05	6.37	6.15	5.89
4.27	3.63	3.55	2.25	3.31	3.73	3.60	6.80	5.97	4.37	4.11
7.27	4.56	4.60	3.92	3.18	4.27	4.48	8.04	7.59	7.41	7.53
2.68	2.23	2.36	1.86	1.91	1.99	1.76	3.82	3.81	2.77	2.68
2.67	2.39	2.54	1.90	1.95	2.20	1.82	3.67	3.82	2.88	2.92
2.50	2.06	2.24	1.72	1.76	1.87	1.39	3.35	3.65	2.72	2.67
2.29	1.71	1.90	1.46	1.60	1.54	0.88	3.10	3.52	2.47	2.36
2.44	2.32	2.40	1.74	1.82	2.21	1.84	3.31	3.27	2.58	2.71
3.05	2.89	2.99	2.01	2.23	2.74	2.56	4.37	3.95	3.16	3.22
3.94	2.40	2.86	1.41	2.66	3.00	2.01	4.48	4.70	2.61	3.01
3.33	4.26	3.95	3.55	2.69	3.08	3.90	5.16	3.83	4.35	3.98
2.62	2.25	2.45	1.93	1.76	1.89	1.71	3.45	3.72	2.88	2.67
2.46	1.99	2.23	1.81	1.62	1.65	1.41	3.17	3.53	2.75	2.48
2.32	1.87	2.06	1.70	1.51	1.52	1.28	3.10	3.35	2.64	2.41
2.37	2.02	2.22	1.73	1.57	1.65	1.50	3.20	3.55	2.68	2.48
2.77	2.34	2.56	2.07	1.84	1.87	1.71	3.38	3.80	3.08	2.83
2.81	2.34	2.48	2.01	1.71	1.83	1.68	3.48	3.50	3.17	2.95
2.73	2.31	2.38	1.78	3.42	1.82	1.63	5.37	3.07	4.97	3.06

Appendix 7.11 Displacement data for the template grip (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2
Phe87 CD2	1.37	2.54	0.67	2.53	2.14	1.94	2.79	2.23	2.99	2.41	2.21
Phe87 CE1	1.69	2.04	0.70	5.18	4.74	4.22	5.37	3.14	2.59	2.85	1.38
Phe87 CE2	1.39	2.64	0.59	2.57	2.19	2.10	2.90	2.48	2.98	2.63	2.17
Phe87 CZ	1.56	0.39	0.57	3.38	2.86	2.34	3.60	1.56	2.76	1.41	1.71
Trp88 N	1.35	0.73	0.40	2.47	2.35	2.29	3.41	1.43	2.27	1.12	1.38
Trp88 CA	1.29	0.62	0.50	2.45	2.36	2.47	3.47	1.55	2.23	1.10	1.42
Trp88 C	1.40	0.69	0.62	2.11	2.13	2.05	2.74	1.29	2.08	1.24	1.51
Trp88 O	1.80	0.73	0.68	2.59	2.29	2.45	3.21	2.55	1.97	2.13	2.05
Trp88 CB	1.20	0.57	0.42	3.06	2.93	3.31	4.35	2.53	2.95	1.51	1.99
Trp88 CG	1.36	0.56	0.51	3.28	3.19	3.61	4.52	2.59	2.98	1.38	2.28
Trp88 CD1	1.40	0.59	0.67	3.45	3.41	3.83	4.65	2.64	3.01	1.19	2.57
Trp88 CD2	1.46	0.49	0.56	3.36	3.24	3.77	4.60	2.62	2.99	1.43	2.34
Trp88 NE1	1.49	0.54	0.88	3.62	3.62	4.10	4.81	2.64	3.02	1.13	2.76
Trp88 CE2	1.51	0.47	0.80	3.56	3.52	4.07	4.78	2.62	3.01	1.28	2.62
Trp88 CE3	1.53	0.46	0.47	3.29	3.10	3.72	4.55	2.70	3.02	1.60	2.28
Trp88 CZ2	1.63	0.40	0.95	3.67	3.64	4.30	4.91	2.67	3.04	1.31	2.76
Trp88 CZ3	1.65	0.38	0.60	3.44	3.26	3.99	4.70	2.83	3.09	1.62	2.50
Trp88 CH2	1.70	0.35	0.84	3.62	3.52	4.26	4.88	2.79	3.10	1.50	2.69
Glu89 N	1.23	0.87	0.66	1.80	2.10	1.73	2.44	0.65	3.77	0.86	1.27
Glu89 CA	1.35	1.37	0.66	1.76	2.35	1.77	2.09	1.00	4.11	1.33	1.72
Glu89 C	1.45	2.23	0.76	2.58	2.85	2.98	2.89	1.78	4.39	2.41	2.37
Glu89 O	1.71	4.34	0.84	2.97	3.13	3.43	3.20	2.19	4.71	2.96	3.44
Glu89 CB	1.14	1.16	0.56	1.63	2.24	1.35	2.30	1.72	4.60	1.40	2.87
Glu89 CG	1.20	0.77	0.30	1.64	0.85	2.36	1.21	3.13	4.89	3.83	3.17
Glu89 CD	1.56	1.72	0.22	2.48	2.38	4.37	3.50	5.31	4.58	5.51	3.43
Glu89 OE1	1.71	2.33	0.25	3.76	3.88	5.01	5.37	7.10	5.44	4.87	5.35
Glu89 OE2	1.74	2.12	0.64	2.74	3.19	5.87	4.59	5.24	4.45	7.72	2.73
Val90 N	1.86	1.25	0.78	3.55	3.63	4.10	3.88	3.38	5.07	4.01	2.24

lrt3	lrt4	lrt5	lrt6	lrt7	lrth	lrti	ltvr	luwb	lvrt	lvru
2.93	2.38	2.53	2.23	2.19	1.80	1.80	2.69	3.69	2.70	2.95
2.74	2.29	2.28	1.70	3.19	1.72	1.55	5.48	2.79	5.07	3.16
2.97	2.38	2.46	2.21	2.45	1.76	1.80	2.78	3.50	2.78	3.07
2.86	2.32	2.31	1.91	1.57	1.69	1.61	3.69	3.03	3.37	3.17
2.21	1.68	1.81	1.61	1.39	1.31	1.05	3.05	3.14	2.56	2.36
2.17	1.64	1.73	1.58	1.35	1.20	0.97	3.15	3.24	2.59	2.46
2.27	1.53	1.47	1.48	1.39	0.89	1.13	2.57	2.43	2.41	2.19
2.19	2.75	2.46	2.62	2.54	2.86	3.24	3.21	2.90	3.02	2.80
2.47	2.11	2.25	2.12	1.81	1.69	1.39	4.05	4.21	3.24	3.21
2.63	1.92	2.02	1.98	1.66	1.38	1.23	4.25	4.39	3.34	3.31
2.79	1.72	1.78	1.79	1.46	1.08	1.03	4.39	4.39	3.42	3.38
2.64	1.93	2.03	2.03	1.71	1.42	1.26	4.32	4.56	3.37	3.37
2.92	1.64	1.63	1.75	1.40	0.97	0.96	4.53	4.50	3.49	3.44
2.82	1.76	1.80	1.90	1.55	1.18	1.10	4.47	4.60	3.45	3.42
2.54	2.10	2.24	2.20	1.89	1.70	1.44	4.26	4.71	3.38	3.45
2.88	1.76	1.80	1.96	1.60	1.25	1.12	4.56	4.74	3.48	3.48
2.62	2.08	2.23	2.24	1.93	1.72	1.45	4.35	4.92	3.46	3.60
2.78	1.93	2.03	2.14	1.80	1.52	1.31	4.50	4.91	3.49	3.58
2.65	1.16	1.04	1.01	0.90	1.64	1.39	2.21	2.21	2.00	1.68
3.09	0.69	0.71	1.31	1.29	0.89	0.97	2.03	1.78	2.06	1.56
3.17	2.98	2.72	2.19	2.27	3.04	3.15	3.35	2.66	3.62	3.28
4.54	4.75	4.09	3.02	3.05	4.39	4.96	4.22	3.21	4.72	4.08
3.91	2.31	1.67	1.80	1.55	1.99	1.89	2.00	2.17	1.50	1.18
4.24	4.02	3.37	4.09	3.88	3.43	3.23	2.36	1.47	3.19	3.53
4.13	5.86	5.80	5.47	5.43	5.73	5.63	4.66	3.84	4.55	5.11
4.92	5.61	6.03	5.07	4.88	6.09	5.70	5.37	5.61	5.49	4.80
4.04	8.07	7.67	7.36	7.52	7.43	7.59	6.12	4.97	6.14	7.10
2.51	4.07	3.93	3.66	3.77	4.34	4.07	4.51	3.81	4.62	4.58

Appendix 7.11 Displacement data for the template grip (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2	1rt3	1rt4	1rt5	1rt6	1rt7	1rth	1rti	1tvr	1uwb	1vrt	1vru
Val90 CA	2.34	1.70	0.91	4.82	4.57	5.51	5.34	4.81	5.93	5.49	3.47	2.74	6.38	5.95	5.11	5.28	6.49	6.16	6.31	5.50	6.30	6.39
Val90 C	3.02	2.47	1.62	5.26	5.31	5.41	5.34	4.96	6.12	5.56	3.50	4.01	6.62	6.03	5.03	5.25	7.04	6.29	6.23	5.57	6.19	6.68
Val90 O	3.74	4.35	3.50	5.95	5.94	6.21	5.94	6.15	7.99	6.37	4.15	5.83	8.17	7.16	5.87	6.10	8.73	8.29	7.24	6.43	6.57	8.21
Val90 CB	2.97	2.38	0.94	6.03	5.51	6.78	6.49	6.53	7.40	7.20	4.38	2.95	8.13	7.52	6.88	7.04	8.16	7.95	7.67	7.04	7.79	7.89
Val90 CG1	3.20	1.23	2.13	5.78	5.03	7.18	6.86	6.66	7.62	7.74	4.03	3.50	8.61	8.06	7.29	7.40	8.89	8.56	8.03	7.34	7.99	8.40
Val90 CG2	3.21	4.22	3.11	6.50	6.15	6.73	6.75	7.06	7.13	7.28	6.12	2.59	8.24	7.98	7.22	7.36	7.94	8.12	8.40	7.14	8.30	7.62
Gln91 N	3.22	1.72	0.27	5.30	5.71	4.74	4.95	3.88	4.40	4.87	3.60	3.17	5.40	5.13	4.17	4.45	5.66	4.32	5.37	4.96	5.78	5.29
Gln91 CA	3.87	2.27	0.77	6.00	6.66	4.70	4.97	4.01	4.95	4.93	4.56	4.30	5.55	5.18	4.15	4.45	6.09	4.28	5.27	5.07	5.72	5.38
Gln91 C	3.44	1.76	0.74	5.12	5.83	4.70	5.22	5.24	5.33	5.84	5.13	4.51	4.89	4.73	5.19	5.41	5.30	5.00	5.37	4.87	6.39	5.62
Gln91 O	3.13	1.30	0.99	4.49	5.28	4.80	5.40	6.22	5.28	6.75	5.54	4.10	4.74	4.65	6.15	6.36	5.10	5.54	5.66	4.74	7.15	6.09
Gln91 CB	4.93	3.67	1.43	7.40	8.17	5.02	5.00	3.39	4.37	4.55	5.62	4.53	6.20	5.58	3.77	4.10	7.06	4.06	5.34	5.78	5.58	5.52
Gln91 CG	5.95	5.06	2.65			6.11	4.47	4.25	6.20	4.26	7.82	7.01	6.72	6.78	3.64	4.06	8.45	3.36	5.23	7.56	6.79	6.27
Gln91 CD	6.76	6.24	3.45			6.59	4.82	6.14	6.93	6.32	9.49	7.63	7.86	9.58	5.77	6.22	11.12	4.38	5.59	10.27	9.79	9.34
Gln91 OE1	7.26	6.09	5.02			6.25	5.17	7.13	9.14	7.85	9.84	9.01	7.78	10.58	7.36	7.79	12.82	5.72	5.68	11.61	10.81	10.66
Gln91 NE2	7.54	7.87	2.93			7.48	5.21	7.44	5.35	6.76	10.79	7.31	9.15	11.20	6.32	6.79	11.88	4.93	6.03	11.44	11.49	10.88
Leu92 N	3.53	1.97	0.63	5.34	5.91	4.98	5.66	5.85	6.62	5.98	6.09	6.24	4.93	4.67	5.64	5.63	5.24	5.82	5.48	5.30	6.57	5.74
Leu92 CA	3.38	1.82	0.63	5.04	5.51	5.35	6.07	7.07	7.35	6.88	6.65	7.28	4.63	4.53	6.69	6.60	4.79	6.61	5.66	5.57	7.37	6.15
Leu92 C	2.56	1.54	0.64	3.97	4.41	4.04	4.61	5.55	5.35	4.93	4.38	5.60	3.40	3.44	4.99	4.85	3.82	4.63	4.20	4.32	5.68	4.48
Leu92 O	2.61	1.81	0.92	3.99	4.32	4.32	4.62	6.52	5.81	5.30	4.14	6.19	3.34	3.32	5.70	5.55	4.27	4.33	5.12	4.60	6.17	4.88
Leu92 CB	4.05	2.31	0.58	6.13	6.47	6.66	7.60	8.69	9.48	8.85	8.85	9.95	5.57	5.22	8.44	8.29	5.35	8.77	7.08	6.82	9.22	8.08
Leu92 CG	4.81	1.19	0.49	6.98	7.92	6.72	7.58	11.30	10.17	11.43	11.52	11.99	5.83	6.29	11.00	10.87	5.20	11.34	7.35	7.00	11.54	10.32
Leu92 CD1	7.21	2.58	0.66	8.01	9.38	9.50	9.96	13.45	12.45	13.26	13.35	14.38	6.43	7.32	13.07	12.84	5.72	13.30	9.82	9.56	13.64	12.36
Leu92 CD2	3.39	2.79	0.41	6.28	7.50	5.59	7.06	11.85	9.79	12.21	11.95	12.66	7.61	7.67	11.60	11.57	6.39	11.72	6.83	6.46	12.25	10.81
Gly93 N	1.94	1.31	0.51	3.42	4.06	2.78	3.61	3.23	3.36	2.89	2.94	3.55	2.71	2.86	2.80	2.66	2.66	3.55	2.76	3.10	3.77	3.24
Gly93 CA	1.27	1.03	0.43	2.33	2.96	1.80	2.47	1.61	1.78	1.09	1.34	1.78	1.58	1.81	1.19	1.07	1.66	1.69	1.71	2.11	2.25	1.75
Gly93 C	0.92	0.90	0.49	2.14	2.70	1.40	1.96	1.17	1.40	0.63	0.75	1.71	1.31	1.44	0.85	0.78	1.29	1.37	1.22	1.61	1.67	1.50
Gly93 O	1.52	1.06	0.74	3.25	3.90	2.48	1.84	1.09	1.54	0.49	0.90	2.00	1.58	1.60	0.79	0.76	1.33	1.37	2.00	1.58	1.77	2.58

Appendix 7.11 Displacement data for the template grip (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2
Val148 N	0.30	0.85	0.24	3.60	3.33	3.12	4.12	2.16	4.83	2.00	2.67
Val148 CA	0.21	0.76	0.28	3.43	3.19	2.96	3.86	2.05	4.62	1.91	2.55
Val148 C	0.30	0.90	0.24	3.53	3.23	2.97	3.98	2.09	4.59	1.90	2.55
Val148 O	0.61	1.35	0.41	3.88	3.36	3.22	4.38	2.16	4.90	1.96	2.68
Val148 CB	0.21	0.51	0.41	3.19	3.11	2.90	3.60	1.88	4.56	1.88	2.46
Val148 CG1	0.16	3.01	2.69	5.81	5.84	5.55	6.15	4.55	7.25	4.63	5.09
Val148 CG2	0.24	2.23	1.86	2.85	3.31	2.34	3.25	2.12	4.01	2.16	1.87
Leu149 N	0.13	0.70	0.37	3.21	3.19	2.72	3.66	2.02	4.23	1.83	2.39
Leu149 CA	0.28	0.71	0.29	3.23	3.20	2.70	3.77	1.99	4.14	1.78	2.31
Leu149 C	0.39	0.70	0.32	3.47	3.29	3.00	4.00	2.11	4.47	1.95	2.55
Leu149 O	0.45	0.52	0.51	3.57	3.09	2.93	3.83	2.08	4.64	1.92	2.67
Leu149 CB	0.36	0.71	0.32	2.96	3.07	2.33	3.45	1.86	3.79	1.67	2.03
Leu149 CG	1.01	0.80	0.27	2.93	3.25	2.19	3.49	1.31	3.66	1.08	1.83
Leu149 CD1	2.67	0.84	0.77	2.77	3.37	2.00	3.43	3.54	3.48	3.35	1.56
Leu149 CD2	2.25	0.97	0.68	3.41	3.83	2.33	3.76	2.93	3.76	2.92	1.84
Pro150 N	0.48	1.01	0.41	3.67	3.64	3.41	4.51	2.36	4.64	2.16	2.72
Pro150 CA	0.52	1.03	0.58	3.93	3.80	3.75	4.80	2.53	4.97	2.40	2.98
Pro150 C	0.76	0.91	0.58	3.84	3.70	3.68	4.73	2.40	4.87	2.30	2.84
Pro150 O	1.08	1.15	0.49	3.73	3.54	3.30	4.37	2.13	4.45	1.98	2.59
Pro150 CB	0.61	1.46	0.56	4.30	4.09	3.77	4.89	2.75	5.06	2.63	3.27
Pro150 CG	0.67	0.47	0.66	4.16	4.13	3.67	4.91	2.74	4.91	2.50	2.07
Pro150 CD	0.76	1.29	0.52	3.75	3.93	3.59	4.93	2.52	4.63	2.20	2.77
Gln151 N	0.58	0.84	0.73	3.94	3.84	4.09	5.12	2.72	5.36	2.72	3.13
Gln151 CA	0.75	0.85	0.84	3.94	3.83	4.22	5.23	2.64	5.38	2.64	3.02
Gln151 C	1.32	1.58	0.92	4.33	4.23	4.08	5.08	3.11	5.68	3.09	3.43
Gln151 O	3.56	3.92	0.89	5.99	5.87	3.77	4.77	5.01	7.17	4.94	5.23
Gln151 CB	0.79	1.22	0.96	3.89	3.86	4.69	5.72	3.06	5.96	3.02	3.57
Gln151 CG	1.81	2.93	3.20	4.49	3.82	5.05	5.85	3.30	5.74	4.23	5.51

lrt3	lrt4	lrt5	lrt6	lrt7	lrth	lrti	ltvr	luwb	lvrt	lvru
4.29	3.19	3.22	2.22	1.98	2.85	2.99	4.60	4.13	4.70	4.67
4.11	3.18	3.16	2.15	1.90	2.82	2.88	4.37	3.81	4.48	4.59
4.08	3.16	3.12	2.12	1.88	2.78	2.85	4.25	3.87	4.51	4.64
4.41	3.19	3.16	2.16	2.00	2.83	2.95	4.15	4.18	4.91	5.09
4.04	3.29	3.27	2.10	1.85	2.98	2.89	4.56	3.56	4.32	4.35
6.75	3.40	3.42	4.74	4.58	3.33	3.09	7.41	6.16	6.97	6.93
2.99	3.16	3.06	2.00	2.13	2.61	2.43	4.03	3.39	3.52	3.58
3.68	3.12	3.07	2.06	1.75	2.73	2.75	4.36	3.60	4.13	4.24
3.60	3.06	3.02	1.98	1.69	2.68	2.70	4.25	3.61	4.11	4.21
3.80	3.33	3.32	2.21	1.86	2.90	2.87	4.44	3.94	4.43	4.45
3.82	3.60	3.65	2.40	1.86	3.09	2.98	4.25	3.90	4.59	4.57
3.28	2.91	2.88	1.81	1.59	2.49	2.52	4.02	3.21	3.72	3.95
3.11	2.85	2.82	1.43	1.09	2.48	2.41	3.87	2.98	3.63	3.87
2.89	2.66	2.64	3.79	3.46	2.31	2.35	3.55	2.72	3.48	3.64
3.12	2.98	2.95	3.03	2.93	2.72	2.34	4.14	3.10	3.82	4.07
4.05	3.40	3.34	2.29	2.05	2.90	3.00	4.94	4.39	4.55	4.68
4.28	3.66	3.61	2.54	2.26	3.17	3.17	5.24	4.81	4.83	4.93
4.10	3.54	3.51	2.45	2.16	3.06	3.08	5.12	4.72	4.68	4.78
3.72	3.24	3.19	2.15	1.84	2.74	2.75	4.96	4.38	4.44	4.53
4.56	3.93	3.86	2.72	2.51	3.52	3.30	5.40	5.10	5.08	5.21
4.56	3.71	3.61	2.54	2.35	2.90	3.29	5.44	5.01	4.86	5.01
4.27	3.38	3.29	2.25	2.09	2.65	3.15	5.33	4.59	4.58	4.78
4.51	3.92	3.93	2.80	2.60	3.51	3.47	5.24	5.08	4.91	4.99
4.44	3.86	3.86	2.76	2.55	3.48	3.45	5.24	5.13	4.83	4.88
4.74	4.34	4.33	3.29	3.08	3.91	3.77	5.08	5.15	5.37	5.25
6.25	6.06	6.02	5.18	4.97	5.61	5.37	4.62	4.90	7.11	6.81
4.93	4.27	4.27	3.15	3.02	4.03	3.94	5.53	5.49	5.05	5.04
5.03	4.05	4.06	4.17	3.94	3.88	4.10	5.78	5.64	5.98	6.80

Appendix 7.11 Displacement data for the template grip (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2
Gln151 CD	0.56	3.44	4.89	3.75	3.58	5.90	6.41	1.36	4.65	2.37	4.29
Gln151 OE1	1.63	3.69	5.01	3.81	3.95	6.29	6.61	2.47	3.49	0.42	2.60
Gln151 NE2	2.61	4.85	7.02	3.78	3.85	6.70	7.06	1.25	5.80	3.36	5.66
Gly152 N	1.11	0.56	1.08	3.49	3.61	4.55	5.50	2.43	5.00	2.47	2.81
Gly152 CA	1.10	0.96	1.13	3.63	3.76	4.45	5.35	2.60	5.12	2.56	3.03
Gly152 C	1.17	1.17	0.94	3.72	3.93	4.21	5.15	2.56	5.04	2.58	2.95
Gly152 O	1.59	2.23	0.92	4.18	4.49	4.16	5.56	2.77	5.40	2.80	3.25
Trp153 N	1.04	0.30	0.89	3.57	3.70	4.20	4.66	2.37	4.66	2.43	2.65
Trp153 CA	1.08	0.31	0.75	3.66	3.88	3.97	4.45	2.24	4.44	2.33	2.42
Trp153 C	1.03	0.44	0.78	3.73	3.98	3.62	4.29	2.16	4.25	2.23	2.33
Trp153 O	1.11	1.06	0.95	4.33	4.82	3.85	4.37	2.41	4.30	2.37	2.58
Trp153 CB	1.19	0.36	0.72	3.60	3.80	3.96	4.50	2.23	4.43	2.27	2.37
Trp153 CG	1.23	0.35	0.56	3.77	3.98	3.80	4.34	2.04	4.18	2.08	2.07
Trp153 CD1	1.31	0.40	0.63	4.11	4.52	3.89	4.60	1.93	4.11	1.95	1.89
Trp153 CD2	1.16	0.40	0.55	3.64	3.63	3.60	4.02	1.94	3.93	1.98	1.94
Trp153 NE1	1.28	0.43	0.78	4.20	4.53	3.80	4.50	1.79	3.81	1.81	1.68
Trp153 CE2	1.19	0.42	0.57	3.92	3.98	3.62	4.15	1.79	3.70	1.82	1.71
Trp153 CE3	1.11	0.51	0.79	3.33	3.13	3.46	3.69	1.97	3.87	2.03	2.01
Trp153 CZ2	1.15	0.51	0.59	3.90	3.79	3.51	3.99	1.70	3.44	1.73	1.58
Trp153 CZ3	1.10	0.61	0.74	3.31	2.99	3.33	3.52	1.88	3.63	1.93	1.88
Trp153 CH2	1.11	0.59	0.51	3.59	3.28	3.36	3.68	1.74	3.42	1.78	1.67
Lys154 N	1.02	0.20	0.71	3.21	3.22	3.31	4.19	1.96	4.14	2.09	2.09
Lys154 CA	1.04	0.34	0.71	3.24	3.23	2.95	4.01	1.89	4.00	2.00	2.01
Lys154 C	1.17	0.36	0.58	3.14	3.03	2.92	3.95	1.73	3.79	1.84	1.87
Lys154 O	1.41	0.69	0.43	3.08	2.91	2.89	3.95	1.65	3.68	1.81	1.85
Lys154 CB	1.08	0.64	0.68	3.37	3.43	2.53	3.57	1.90	4.05	1.99	2.00
Lys154 CG	1.12	1.95	0.99	3.90	3.70	2.41	3.34	1.86	3.86	1.98	2.13
Lys154 CD	1.32	1.42	1.13	2.49	2.08	1.72	2.67	1.86	3.75	2.00	2.13

lrt3	lrt4	lrt5	lrt6	lrt7	lrth	lrti	ltvr	luwb	lvrt	lvru
4.97	3.55	3.70	2.18	1.75	3.48	3.94	4.97	6.31	6.69	6.50
3.98	3.68	3.77	0.70	0.72	3.61	3.20	4.76	6.58	5.94	4.81
6.72	3.23	3.47	3.06	2.27	3.20	5.47	5.05	7.05	8.33	8.51
4.16	3.62	3.70	2.59	2.44	3.34	3.30	5.71	5.51	4.47	4.51
4.24	3.84	3.92	2.81	2.62	3.50	3.32	5.61	5.56	4.85	4.65
4.19	3.74	3.76	2.81	2.63	3.33	3.24	5.83	5.57	4.85	4.82
4.45	3.86	3.79	2.97	2.76	3.34	3.33	7.39	6.53	5.27	5.51
3.95	3.56	3.60	2.68	2.53	3.19	3.10	4.56	4.66	4.46	4.41
3.81	3.39	3.38	2.57	2.41	2.97	2.90	4.62	4.59	4.29	4.32
3.63	3.33	3.35	2.56	2.43	2.91	2.79	4.66	4.36	4.08	4.14
3.70	3.71	3.84	2.85	2.81	3.19	3.10	5.38	4.59	4.13	4.25
3.83	3.34	3.32	2.55	2.35	2.88	2.83	4.58	4.63	4.27	4.30
3.66	3.12	3.08	2.33	2.11	2.68	2.56	4.72	4.48	4.01	4.05
3.60	2.99	2.95	2.23	1.88	2.55	2.44	5.26	4.48	3.91	3.99
3.51	3.01	2.97	2.15	2.04	2.55	2.39	4.40	4.17	3.83	3.85
3.42	2.83	2.78	2.03	1.69	2.38	2.20	5.31	4.20	3.69	3.77
3.37	2.83	2.79	1.98	1.78	2.37	2.17	4.81	4.02	3.64	3.69
3.49	3.05	3.02	2.14	2.18	2.58	2.40	3.84	4.03	3.82	3.81
3.23	2.71	2.67	1.83	1.67	2.24	1.98	4.68	3.78	3.47	3.52
3.37	2.93	2.90	1.97	2.07	2.44	2.23	3.72	3.77	3.67	3.66
3.24	2.76	2.73	1.83	1.81	2.27	2.02	4.15	3.66	3.49	3.51
3.53	2.96	2.94	2.35	2.20	2.62	2.50	3.90	4.11	3.94	3.91
3.40	2.86	2.85	2.31	2.16	2.53	2.38	3.78	3.90	3.79	3.75
3.24	2.78	2.79	2.18	2.07	2.42	2.24	3.62	3.72	3.58	3.53
3.16	2.78	2.78	2.11	2.01	2.36	2.22	3.56	3.60	3.50	3.38
3.47	2.74	2.73	2.34	2.15	2.46	2.36	3.69	3.78	3.86	3.77
3.50	2.77	2.72	2.31	2.07	2.54	2.40	3.66	3.46	3.57	3.72
3.49	2.65	2.61	2.34	2.11	2.32	2.30	3.03	3.00	3.61	3.54

Appendix 7.11 Displacement data for the template grip (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2	1rt3	1rt4	1rt5	1rt6	1rt7	1rth	1rti	1tvr	1uwb	1vrt	1vru
Lys154 CE	2.00	1.73	1.97	3.24	3.12	3.64	4.78	1.96	3.64	2.09	2.24	3.54	2.76	2.69	2.43	2.29	2.39	2.36	4.87	5.01	3.39	3.60
Lys154 NZ	3.53	1.08	1.37	3.63	3.26	2.88	4.02	2.01	3.40	2.18	2.25	3.44	2.61	2.64	2.50	2.53	2.40	2.22	4.24	4.43	3.44	3.49

The measured distances between the unliganded (1dlo) structure and each RT structure for all the atoms of the residues of template grip.

Appendix 7.12 Displacement data for the RNase H active site

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1
Asp443 N	0.26	0.55	0.35	3.48	3.38	3.76	4.25	2.73	1.56	3.25
Asp443 CA	0.11	0.43	0.30	3.59	3.49	3.83	4.33	2.81	1.62	3.33
Asp443 C	0.39	0.48	0.50	3.63	3.61	3.90	4.49	2.89	1.63	3.41
Asp443 O	0.54	0.60	1.01	3.65	3.65	3.89	4.48	3.02	1.51	3.50
Asp443 CB	0.41	0.41	0.58	3.73	3.65	3.83	4.23	2.83	1.74	3.25
Asp443 CG	0.78	0.45	0.67	3.79	3.68	3.88	4.23	2.85	2.73	3.36
Asp443 OD1	0.81	1.35	0.74	3.71	3.62	4.21	4.49	2.93	2.96	3.31
Asp443 OD2	1.17	2.20	0.85	3.91	3.76	3.68	4.42	2.91	3.75	3.65
Glu478 N	0.84	1.11	0.82	3.63	3.85	4.14	4.65	2.84	2.17	3.31
Glu478 CA	0.81	0.77	0.56	3.50	3.80	3.95	4.41	2.62	2.20	3.03
Glu478 C	0.77	0.78	0.43	3.48	3.56	3.96	4.47	2.58	2.28	2.98
Glu478 O	0.99	0.94	0.40	3.48	3.41	4.04	4.61	2.92	2.45	3.20
Glu478 CB	0.83	0.73	0.49	3.47	3.77	3.90	4.34	2.45	2.24	2.83
Glu478 CG	0.71	0.91	0.41	3.02	4.44	3.62	3.88	1.79	1.98	2.24
Glu478 CD	2.20	1.42	1.08	2.54	2.88	2.34	2.46	1.36	2.52	1.15
Glu478 OE1	3.87	1.58	1.33	3.57	1.67	3.09	2.96	2.19	3.77	3.17
Glu478 OE2		1.73	1.47	2.87	3.91	2.36	2.49	3.35	3.04	1.96
Asp498 N	0.58	0.26	0.38	3.42	3.57	3.08	3.40	1.93	1.63	2.21
Asp498 CA	0.51	0.22	0.53	3.68	3.84	3.10	3.42	1.79	1.57	2.05
Asp498 C	0.47	0.06	0.57	3.43	3.56	3.10	3.44	1.85	1.64	2.26
Asp498 O	0.54	0.53	1.54	3.75	3.73	3.11	3.48	1.81	1.67	2.31
Asp498 CB	0.53	0.15	0.65	4.68	4.83	3.34	3.72	1.76	1.53	1.99
Asp498 CG	0.56	0.10	0.64	3.43	3.63	3.78	4.11	2.20	1.87	2.50
Asp498 OD1	0.68	0.47	0.59	2.91	3.08	4.11	4.27	1.22	1.88	1.49
Asp498 OD2	0.55	0.37	0.88	4.15	4.33	4.10	4.35	3.93	2.14	4.09
His539 N	1.11	1.83	1.43	2.18	2.37	3.45	3.81	2.83	3.00	3.16
His539 CA	1.53	1.97	1.79	1.94	2.55	3.45	3.83	3.17	3.32	3.51
His539 C	1.50	1.81	1.83	1.78	2.52	2.77	3.28	3.47	3.14	3.79
His539 O	1.61	1.72	2.06	1.88	3.42	2.70	3.40	3.66	3.76	5.64

lrt2	lrt3	lrt4	lrt5	lrt6	lrt7	lrth	lrti	ltvr	luwb	lvrt	lvru
2.63	1.59	2.70	2.55	2.86	3.07	3.03	2.71	3.46	4.17	1.74	1.82
2.73	1.58	2.73	2.57	2.93	3.16	3.07	2.76	3.54	4.24	1.79	1.96
2.69	1.58	2.78	2.61	2.99	3.23	3.09	2.88	3.83	4.44	1.71	2.28
2.76	1.54	2.90	2.72	3.06	3.35	3.14	3.02	3.93	4.52	1.65	2.43
2.83	1.50	2.69	2.51	2.86	3.16	2.98	2.77	3.38	4.17	1.84	1.98
2.96	3.18	2.68	2.46	2.98	3.29	2.94	2.75	3.36	4.20	3.55	3.34
2.93	5.56	3.03	2.77	2.97	3.22	3.23	2.94	3.50	4.37	4.24	3.88
3.52	2.50	2.56	2.40	3.14	3.50	2.85	2.98	3.28	4.31	4.57	4.37
2.49	1.95	2.74	2.53	2.92	3.25	2.94	2.77	3.80	4.42	2.17	1.93
2.15	1.99	2.58	2.37	2.75	3.06	2.83	2.64	3.81	4.34	2.21	1.86
2.21	2.11	2.59	2.40	2.69	3.05	2.82	2.58	3.74	4.40	2.40	2.02
2.43	2.34	3.00	2.76	2.96	3.32	3.10	2.95	3.53	4.67	2.77	2.27
1.90	2.07	2.45	2.23	2.61	2.90	2.74	2.58	3.85	4.30	2.30	1.89
1.38	1.84	1.73	1.58	1.95	2.20	2.07	2.05	3.23	3.72	1.96	1.53
1.59	2.81	1.09	1.29	1.10	1.15	0.88	0.67	2.56	2.07	2.60	2.80
2.44	4.18	2.58	2.50	3.04	1.83	2.04	2.23	3.30	2.38	3.58	3.79
3.11	3.26	1.63	2.09	1.83	3.22	1.63	0.99	3.13	2.31	3.33	3.93
1.79	1.59	2.14	1.97	1.79	2.11	2.29	2.11	3.17	3.16	1.81	1.75
1.60	1.54	1.99	1.83	1.60	1.93	2.15	1.90	3.33	3.13	1.73	1.69
1.79	1.66	2.04	1.88	1.73	2.06	2.25	2.02	3.12	3.23	1.86	1.78
1.91	1.76	2.01	1.84	1.70	2.03	2.29	1.99	3.41	3.40	1.89	1.83
1.36	1.47	2.03	1.86	1.50	1.83	2.13	1.82	3.85	3.43	1.63	1.65
1.81	1.80	2.34	2.16	1.70	2.21	2.40	2.22	4.13	3.96	2.14	2.11
0.72	2.13	2.52	2.34	0.77	1.10	2.23	3.21	4.35	4.07	2.45	2.58
3.38	1.92	2.41	2.23	3.34	3.93	2.81	1.97	4.35	4.33	2.27	2.16
1.92		2.68	2.95	3.43	3.13	3.06	2.51	3.63	3.01	1.73	2.04
1.97		3.06	3.44	4.17	3.52	3.47	2.81	3.75	3.39	2.69	2.95
2.37		2.99	3.46	4.57	3.84	3.26	2.90	2.93	2.91	3.10	3.41
2.69		3.14	3.67	6.42	5.72	3.34	3.20	2.74	3.37	3.70	4.13

Appendix 7.12 Displacement data for the RNase H active site (cont.).

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2	1rt3	1rt4	1rt5	1rt6	1rt7	1rth	1rti	1tvr	1uwb	1vrt	1vru
His539 CB	1.94	2.34	1.83	2.24	2.62	4.02	4.16	3.20	3.46	3.58	1.96		3.23	3.65	4.41	3.61	3.69	2.85	4.28	3.75	3.23	3.34
His539 CG	2.24	2.72	1.98	2.29	2.77	4.79	4.80	2.98	4.50	3.32	1.70		3.39	3.81	4.13	3.36	3.81	2.84	4.69	4.26	2.99	2.94
His539 ND1	4.37	3.25	1.88	3.84	4.84	5.88	6.17	4.88	6.21	5.18	3.44		5.42	5.96	5.87	5.19	5.74	4.71	4.89	5.12	3.14	3.26
His539 CD2	1.53	3.94	2.42	1.97	1.76	4.86	4.52	1.89	4.90	2.15	1.68		2.09	2.41	2.93	2.24	2.50	1.79	4.97	4.18	2.92	2.45
His539 CE1	4.06	3.67	2.08	3.56	4.51	6.44	6.49	4.15	6.90	4.41	2.66		5.06	5.58	5.09	4.44	5.28	4.24	5.26	5.38	2.91	2.94
His539 NE2	1.91	3.87	2.46	1.99	2.39	5.85	5.53	2.16	6.23	2.47	1.35		2.93	3.33	3.27	2.53	3.22	2.25	5.30	4.82	2.84	2.35
Asp549 N	1.17	1.32	0.80	3.33	3.25	3.85	4.03												3.15	3.52		
Asp549 CA	1.06	1.74	0.97	3.75	3.64	4.09	4.12												3.16	3.62		
Asp549 C	1.13	1.73	0.96	3.88	3.88	4.23	4.52												3.47	4.04		
Asp549 O	1.04	1.78	0.92	3.65	3.83	4.12	4.63												3.31	4.07		
Asp549 CB	1.21	2.20	1.44	3.54	3.46	3.96	4.09												3.13	3.50		

The measured distances between the unliganded (1dlo) structure and each RT structure for all the atoms of the residues of the RNase H active site.

Appendix 7.13 Displacement data for the NNIBP

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2
Leu100 N	0.42	0.31	1.16	1.83	1.82	1.20	1.86	1.31	1.48	1.41	1.34
Leu100 CA	0.39	0.35	1.11	2.09	2.13	1.35	1.95	1.41	1.56	1.48	1.47
Leu100 C	0.30	0.31	1.11	1.98	2.02	1.37	1.89	1.13	1.56	1.28	1.33
Leu100 O	0.60	0.44	1.63	1.82	1.67	1.51	1.70	1.01	1.57	1.20	1.19
Leu100 CB	0.42	0.48	1.10	2.39	2.52	1.70	2.06	1.57	1.69	1.56	1.53
Leu100 CG	0.64	2.49	1.12	2.31	2.62	2.09	2.29	1.56	1.92	1.87	1.69
Leu100 CD1	0.75	4.13	1.13	2.47	2.81	2.34	2.30	1.74	1.91	1.61	1.51
Leu100 CD2	0.73	2.80	1.15	2.16	2.81	2.90	2.81	1.40	2.33	2.31	1.99
Lys101 N	0.12	0.52	0.76	2.19	2.43	1.39	2.08	1.39	1.63	1.43	1.67
Lys101 CA	0.23	0.52	0.73	2.23	2.50	1.59	2.12	1.32	1.64	1.36	1.62
Lys101 C	0.31	0.71	0.73	2.28	2.65	1.89	2.26	1.26	1.70	1.44	1.60
Lys101 O	0.53	0.67	0.86	2.42	3.04	1.77	2.38	1.11	1.84	1.36	1.50
Lys101 CB	0.41	0.66	0.92	2.36	2.58	1.58	2.08	1.59	1.66	1.61	1.76
Lys101 CG	0.56	1.96	1.14	2.19	2.11	1.40	2.09	1.21	1.71	1.90	1.56
Lys101 CD	1.47	2.31	1.36	1.79	2.13	1.40	1.73	1.36	3.62	1.56	3.15
Lys101 CE	0.94	1.94	1.70	1.51	2.22	1.48	1.90	2.32	3.46	1.86	2.32
Lys101 NZ	1.35	3.13	2.12	1.41	2.02	1.56	1.89	1.97	5.11	1.09	3.35
Lys103 N	0.63	1.56	0.86	2.52	2.62	2.93	2.64	1.45	2.15	1.69	1.64
Lys103 CA	0.76	1.86	1.01	2.61	2.70	3.11	2.76	1.38	2.45	1.68	1.65
Lys103 C	1.05	2.20	1.13	2.78	3.03	3.32	3.02	1.42	2.55	1.64	1.60
Lys103 O	2.03	3.29	2.07	2.60	3.05	3.71	3.46	1.87	2.99	2.18	2.06
Lys103 CB	0.93	2.00	1.22	2.43	2.40	3.12	2.77	1.29	2.66	1.74	1.69
Lys103 CG	0.82	1.74	1.25	1.77	1.73	3.31	2.79	1.22	2.56	2.07	1.69
Lys103 CD	0.92	1.80	1.26	2.34	2.49	3.21	2.72	1.18	2.58	2.11	1.63
Lys103 CE	1.03	1.61	1.26	1.32	1.71	3.39	2.69	1.04	2.41	2.42	1.54
Lys103 NZ	1.56	1.72	1.28	2.59	2.84	3.37	2.69	1.15	4.05	2.47	1.54
Val106 N	0.22	1.41	0.48	3.20	3.23	3.50	3.41	1.28	2.65	2.14	2.07
Val106 CA	0.16	1.44	0.53	2.96	2.94	3.13	3.12	1.14	2.45	1.99	1.86
Val106 C	0.23	1.40	0.53	3.06	3.17	3.35	3.10	1.07	2.56	1.74	1.60

lrt3	lrt4	lrt5	lrt6	lrt7	lrth	lrti	ltvr	luwb	lvrt	lvru
1.28	1.33	1.24	1.37	1.63	1.08	1.46	1.88	2.18	1.26	1.18
1.46	1.39	1.30	1.44	1.71	1.12	1.50	1.96	2.26	1.35	1.19
1.40	1.18	1.11	1.26	1.52	1.02	1.32	1.90	2.15	1.25	1.18
1.43	0.94	0.97	1.10	1.33	0.91	1.22	2.06	1.97	1.17	1.18
1.68	1.50	1.42	1.56	1.79	1.20	1.48	2.15	2.41	1.46	1.23
1.92	1.71	1.67	1.86	2.21	1.55	1.72	2.25	2.67	1.75	1.42
2.04	1.56	1.47	1.83	1.86	1.47	1.58	2.32	2.51	1.68	1.44
2.20	1.99	2.03	2.23	3.10	2.06	2.03	2.55	3.27	2.23	1.91
1.47	1.39	1.23	1.37	1.64	1.08	1.45	1.88	2.33	1.42	1.22
1.48	1.32	1.17	1.30	1.59	1.09	1.40	1.84	2.32	1.46	1.28
1.49	1.37	1.32	1.40	1.60	1.22	1.28	1.98	2.34	1.61	1.55
1.55	1.35	1.51	1.48	1.60	1.17	1.15	2.00	2.30	1.50	1.55
1.53	1.46	1.31	1.37	1.68	1.07	1.72	1.77	2.31	1.56	1.57
1.38	1.61	0.87	1.30	1.43	1.07	1.99	1.65	2.29	1.33	1.29
3.22	1.99	1.54	3.16	3.07	1.06	2.25	1.42	2.03	3.19	2.83
2.74	2.48	2.51	2.89	2.70	1.18	2.26	1.21	2.01	2.91	2.08
4.31	4.49	4.55	4.81	4.71	2.96	2.19	1.06	1.49	4.65	3.82
1.99	1.88	1.84	1.63	1.76	1.83	1.34	2.42	2.67	2.41	2.65
2.24	2.05	2.05	1.73	1.88	2.02	1.36	2.38	2.86	2.66	2.97
2.43	2.12	2.09	1.63	1.88	2.08	1.56	2.68	3.25	2.78	2.92
3.00	2.63	2.61	2.22	2.60	2.61	2.26	2.91	3.72	3.27	3.27
2.41	2.25	2.27	1.99	2.04	2.32	1.40	2.22	2.76	3.01	3.52
2.24	2.35	2.33	2.21	2.23	2.31	1.63	2.18	2.87	3.04	3.68
2.29	2.29	2.26	2.15	2.24	2.28	1.65	2.12	2.88	3.06	3.58
2.31	2.35	2.25	2.36	2.46	2.19	1.90	2.03	3.06	2.87	3.50
2.52	2.50	2.44	2.64	3.07	2.56	1.70	2.40	3.19	3.10	3.65
2.33	2.19	2.33	1.91	2.06	2.34	1.60	3.52	3.39	3.03	3.01
2.10	1.90	2.08	1.72	1.82	2.08	1.30	3.15	3.01	2.73	2.72
2.11	1.93	2.13	1.84	1.84	2.02	1.16	3.40	3.06	2.80	2.75

Appendix 7.13 Displacement data for the NNIBP (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2
Val106 O	0.29	1.42	0.78	3.40	3.92	3.84	3.00	1.17	3.02	1.99	1.77
Val106 CB	0.09	1.46	0.38	2.98	2.81	2.93	3.24	1.30	2.55	2.37	2.38
Val106 CG1	0.08	1.57	0.33	2.70	2.78	2.88	3.17	1.32	2.45	2.40	2.43
Val106 CG2	0.20	1.55	0.76	2.76	2.63	2.81	3.18	1.15	2.34	2.48	2.56
Val179 N	0.32	1.06	1.71	2.02	2.72	2.20	2.54	0.49	1.47	0.69	0.57
Val179 CA	0.23	1.19	1.41	1.95	2.35	2.16	2.28	0.37	1.34	0.66	0.51
Val179 C	0.24	1.13	1.05	1.74	2.04	2.00	2.12	0.46	1.28	0.69	0.54
Val179 O	0.28	1.09	0.83	1.74	1.87	2.08	2.30	0.58	1.38	0.66	0.63
Val179 CB	0.72	1.47	1.72	1.82	2.07	2.36	2.18	0.30	1.46	0.62	0.51
Val179 CG1	1.61	0.52	2.03	3.29	3.18	3.82	3.67	2.18	1.59	2.85	2.45
Val179 CG2	1.99	2.83	2.15	1.67	2.10	0.80	0.86	1.66	2.16	1.92	1.56
Tyr181 N	0.97	1.18	0.60	1.75	1.79	1.70	1.97	0.83	1.50	1.00	0.78
Tyr181 CA	0.99	1.02	0.56	1.83	1.81	1.76	2.13	0.95	1.66	1.12	0.87
Tyr181 C	0.96	1.01	0.74	1.85	1.80	1.78	2.10	0.94	1.67	1.12	0.79
Tyr181 O	1.17	1.13	0.90	1.89	1.79	1.80	1.91	0.93	1.59	1.21	0.84
Tyr181 CB	0.94	0.96	0.34	1.81	1.70	1.58	2.19	0.92	1.58	1.04	0.93
Tyr181 CG	0.98	1.00	1.02	3.75	3.77	3.58	4.00	2.96	3.59	3.07	2.95
Tyr181 CD1	0.80	1.39	1.71	4.97	6.31	5.05	5.19	4.73	5.09	4.92	4.80
Tyr181 CD2	1.25	1.80	2.61	5.09	3.28	4.63	5.04	4.02	4.70	4.00	3.87
Tyr181 CE1	0.82	1.33	2.19	6.85	8.31	6.84	6.87	6.53	6.91	6.79	6.61
Tyr181 CE2	1.31	1.97	3.18	6.85	5.23	6.35	6.65	5.79	6.50	5.79	5.65
Tyr181 CZ	1.06	1.19	2.48	7.56	7.74	7.27	7.41	6.81	7.40	6.98	6.80
Tyr181 OH	1.08	1.38	3.23	9.44	9.69	9.08	9.10	8.70	9.28	8.95	8.69
Tyr188 N	0.92	0.99	0.82	1.86	2.01	2.08	2.38	0.48	1.86	0.55	0.56
Tyr188 CA	0.89	0.94	0.80	2.02	2.12	2.27	2.53	0.70	1.89	0.79	0.82
Tyr188 C	0.90	1.12	0.79	1.93	2.07	2.14	2.33	0.71	1.60	0.70	0.77
Tyr188 O	1.41	1.22	0.83	2.06	1.80	2.04	2.03	1.32	1.37	1.25	0.96
Tyr188 CB	1.00	0.80	0.69	2.26	2.15	2.50	2.70	1.13	2.09	1.17	1.14
Tyr188 CG	1.02	0.87	0.65	4.74		4.94	5.10	3.75	4.40	3.74	3.74

lrt3	lrt4	lrt5	lrt6	lrt7	lrth	lrti	ltvr	luwb	lvrt	lvru
2.58	2.45	2.78	2.58	2.40	2.58	1.35	4.60	3.01	3.49	3.28
2.26	2.06	2.19	1.83	1.98	2.25	1.65	3.18	3.14	2.79	2.65
2.10	1.87	1.99	1.79	1.93	2.10	1.63	3.09	2.75	2.58	2.12
2.10	1.90	1.99	1.57	1.82	2.15	1.61	2.76	2.97	2.65	2.62
1.63	1.52	1.66	1.04	0.82	1.82	0.35	2.12	1.78	2.18	2.50
1.30	1.40	1.51	0.76	0.72	1.66	0.25	2.05	1.59	1.88	2.22
1.09	1.27	1.38	0.67	0.72	1.52	0.27	1.97	1.54	1.72	2.05
1.13	1.31	1.43	0.83	0.72	1.71	0.23	2.12	1.69	1.77	2.08
1.41	1.53	1.60	0.63	0.68	1.86	0.33	2.04	1.30	1.93	2.43
2.54	2.88	2.96	1.91	2.25	3.15	2.86	3.80	2.80	3.07	3.57
0.83	0.99	0.98	1.72	1.35	0.97	2.41	1.14	1.55	0.97	0.92
0.69	1.29	1.41	1.03	1.03	1.41	0.59	2.14	1.81	1.52	1.76
0.63	1.32	1.45	1.08	1.24	1.44	0.69	2.22	2.01	1.55	1.79
0.55	1.34	1.45	1.07	1.24	1.47	0.60	2.15	1.89	1.56	1.84
0.63	1.28	1.34	1.07	1.29	1.36	0.52	1.97	1.62	1.56	1.80
0.48	1.12	1.25	1.05	1.23	1.18	0.80	2.07	2.03	1.52	1.59
0.83	3.01	3.13	3.09	3.28	3.10	0.74	3.89		3.39	3.41
2.22	4.70	4.64	4.89	4.76	4.72	0.70	6.36		4.70	4.67
2.20	4.08	4.31	4.02	4.46	4.34	1.28	3.29		4.60	4.78
2.39	6.54	6.49	6.70	6.62	6.59	0.64	8.14		6.56	6.62
2.59	5.76	6.04	5.79	6.19	6.10	1.39	4.96		6.27	6.51
1.81	6.79	6.93	6.90	7.08	7.00	0.86	7.39		7.09	7.28
2.33	8.66	8.81	8.78	8.97	8.97	1.00	9.13		8.91	9.27
1.27	1.34	1.50	0.78	0.76	1.58	0.61	1.82	2.03	1.47	1.88
1.46	1.36	1.54	0.95	0.89	1.65	0.72	1.96	2.16	1.76	2.03
1.37	1.13	1.29	0.75	0.74	1.35	0.50	1.83	1.99	1.63	1.95
1.30	1.00	1.15	1.00	1.11	1.16	0.57	1.62	1.69	1.51	1.86
1.69	1.62	1.77	1.35	1.30	1.91	1.10	2.17	2.37	2.05	2.19
4.32	4.05	4.19	3.99	3.93	4.24	3.55	4.70	4.73	4.46	4.36

Appendix 7.13 Displacement data for the NNIBP (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2
Tyr188 CD1	1.07	0.80	0.76	7.26		7.38	7.61	6.02	6.75	5.90	5.90
Tyr188 CD2	1.02	1.08	0.55	4.63		4.85	4.91	4.01	4.34	4.11	4.14
Tyr188 CE1	1.10	0.85	0.69	9.56		9.73	9.87	8.50	9.00	8.39	8.39
Tyr188 CE2	1.05	1.21	0.46	6.93		7.21	7.18	6.49	6.53	6.58	6.62
Tyr188 CZ	1.10	1.08	0.49	9.36		9.63	9.65	8.71	8.87	8.71	8.73
Tyr188 OH	1.17	1.24	0.35	11.64		11.95	11.95	11.17	11.09	11.18	11.23
Gly190 N	0.47	1.37	0.65	2.45	2.36	2.63	2.69	1.14	1.72	1.14	1.00
Gly190 CA	0.44	1.46	0.56	2.70	2.35	2.83	2.96	1.17	1.97	1.25	0.98
Gly190 C	0.23	1.28	0.73	2.72	2.42	2.90	2.99	1.19	2.00	1.18	0.95
Gly190 O	0.40	1.09	1.01	2.59	2.34	2.90	2.83	1.29	1.87	1.03	0.88
Phe227 N	0.92	1.62	0.75	4.47	4.50	4.32	4.58	2.09	3.51	3.08	2.72
Phe227 CA	0.88	1.50	0.85	4.94	4.85	4.59	4.90	2.70	3.94	3.67	3.20
Phe227 C	0.91	1.45	0.70	4.88	4.80	4.47	4.80	2.73	3.83	3.67	3.10
Phe227 O	0.76	1.42	0.72	4.67	4.59	4.12	4.55	2.46	3.66	3.32	2.84
Phe227 CB	0.84	1.36	0.95	5.10	4.77	4.57	5.05	2.87	4.22	3.90	3.58
Phe227 CG	0.67	1.55	0.92	5.48	6.26	4.63	5.12	2.93	4.47	4.10	3.70
Phe227 CD1	0.87	1.78	1.00	6.09	6.01	4.87	5.28	3.12	4.76	4.45	3.70
Phe227 CD2	0.39	1.56	0.82	5.35	8.36	4.49	5.05	2.91	4.53	4.12	3.87
Phe227 CE1	0.75	1.96	0.97	6.51	7.60	4.99	5.36	3.19	5.01	4.70	3.80
Phe227 CE2	0.26	1.80	0.78	5.85	10.12	4.58	5.11	2.98	4.83	4.44	4.02
Phe227 CZ	0.42	1.97	0.85	6.39	9.67	4.84	5.26	3.07	5.01	4.67	3.94
Trp229 N	0.99	1.21	0.81	4.59	4.22	4.71	5.12	4.67	4.44	5.14	3.89
Trp229 CA	0.97	1.13	0.81	4.65	4.21	4.88	5.36	4.11	4.85	5.71	4.32
Trp229 C	1.19	1.23	0.80	4.68	4.27	4.73	5.07	4.41	4.82	5.93	4.54
Trp229 O	1.45	1.35	0.82	4.65	4.44	4.36	4.75	4.23	4.59	5.72	4.38
Trp229 CB	0.85	1.10	0.77	4.58	4.29	4.77	5.32	4.10	4.80	5.71	4.33
Trp229 CG	0.99	1.07	0.75	4.77	4.52	4.98	5.44	4.19	5.21	5.25	4.50
Trp229 CD1	1.01	1.08	0.90	5.53	5.45	5.46	5.63	4.51	5.59	5.36	4.56
Trp229 CD2	1.20	1.14	0.74	5.14	4.77	4.72	5.45	4.05	5.26	5.06	4.60

lrt3	lrt4	lrt5	lrt6	lrt7	lrth	lrti	ltvr	luwb	lvrt	lvru
6.50	6.34	6.47	6.23	6.20	6.44	5.78	7.02	7.25	6.61	6.50
5.48	4.19	4.33	4.24	4.20	4.50	3.83	4.80	4.51	4.71	4.59
8.86	8.75	8.87	8.69	8.68	8.86	8.18	9.43	9.49	8.96	8.87
7.92	6.57	6.71	6.68	6.66	6.85	6.18	7.20	6.72	6.99	6.89
9.34	8.84	8.98	8.89	8.89	9.03	8.33	9.51	9.20	9.11	9.05
11.85	11.23	11.38	11.31	11.35	11.43	10.77	11.88	11.40	11.43	11.42
1.57	1.26	1.36	1.32	1.15	1.35	0.85	2.06	2.57	1.91	2.26
1.65	1.34	1.45	1.46	1.26	1.47	0.98	2.31	2.84	2.01	2.45
1.63	1.43	1.50	1.39	1.26	1.55	1.01	2.34	2.89	2.18	2.57
1.49	1.55	1.53	1.27	1.21	1.56	0.91	2.45	2.78	2.15	2.53
2.89	2.62	2.64	2.89	3.06	2.61	2.22	5.50	5.07	3.43	3.35
3.48	3.17	3.30	3.48	3.65	3.08	2.73	5.77	5.32	3.89	3.82
3.42	3.15	3.26	3.48	3.68	2.98	2.77	5.71	5.26	3.72	3.63
3.08	2.79	2.92	3.12	3.34	2.66	2.55	5.47	5.06	3.32	3.23
3.85	3.39	3.59	3.66	3.83	3.35	3.07	5.61	5.27	4.16	4.12
3.98	3.76	3.95	3.95	4.03	3.92	3.26	5.75	5.41	4.38	4.31
4.18	4.24	4.56	4.47	4.53	4.68	3.53	6.23	5.78	4.70	4.67
4.10	3.94	3.90	3.94	3.92	3.93	3.42	5.50	5.22	4.49	4.25
4.32	4.67	4.99	4.81	4.80	5.28	3.76	6.45	5.95	4.95	4.87
4.30	4.52	4.44	4.35	4.25	4.64	3.78	5.70	5.38	4.78	4.52
4.32	4.78	4.92	4.71	4.63	5.23	3.85	6.18	5.73	4.93	4.77
4.06	4.40	4.53	4.56	4.89	3.74	3.68	5.47	5.23	4.34	4.27
4.32	4.85	4.99	5.04	5.43	4.05	4.08	5.75	5.43	4.67	4.62
4.36	5.08	5.21	5.29	5.72	4.20	4.28	5.36	5.30	4.57	4.46
4.16	4.97	5.05	5.12	5.57	4.01	4.17	4.99	5.04	4.39	4.17
4.18	4.89	4.98	5.12	5.55	4.00	4.02	5.87	5.39	4.57	4.55
4.32	5.10	5.10	5.34	5.55	4.15	4.46	5.83	5.59	4.82	4.87
4.24	6.59	6.59	6.91	6.90	4.22	4.78	6.15	5.69	5.01	5.11
4.55	4.51	4.36	4.55	4.81	4.14	4.66	5.49	5.70	4.85	4.97

Appendix 7.13 Displacement data for the NNIBP (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2
Trp229 NE1	1.17	1.16	1.00	5.54	5.28	5.49	5.76	4.57	5.85	5.26	4.67
Trp229 CE2	1.26	1.20	0.88	4.98	4.64	5.06	5.66	4.29	5.67	5.08	4.70
Trp229 CE3	1.38	1.24	0.73	6.58	6.19	4.24	5.31	3.77	5.00	5.15	4.64
Trp229 CZ2	1.46	1.33	0.93	5.26	4.98	4.94	5.74	4.29	5.82	5.22	4.85
Trp229 CZ3	1.58	1.36	0.82	7.48	7.27	4.11	5.38	3.75	5.15	5.26	4.77
Trp229 CH2	1.61	1.41	0.89	6.50	6.38	4.46	5.60	4.01	5.56	5.30	4.88
Leu234 N	0.96	1.32	1.06	4.03	3.94	3.40	3.63	1.69	2.73	2.17	1.99
Leu234 CA	0.90	1.36	1.08	3.87	3.76	3.32	3.48	1.47	2.70	1.93	1.89
Leu234 C	0.80	1.38	1.07	3.97	3.81	3.34	3.63	1.11	2.70	1.76	1.85
Leu234 O	0.81	1.11	1.07	4.73	4.68	4.15	4.44	1.38	3.26	2.48	2.32
Leu234 CB	0.87	1.30	1.08	4.13	3.96	3.71	3.75	1.74	3.19	2.30	2.23
Leu234 CG	0.84	1.67	1.18	3.71	4.03	3.49	3.69	2.04	2.81	2.53	2.17
Leu234 CD1	0.97	1.82	0.86	4.40	5.11	3.85	4.84	3.98	3.97	5.06	4.34
Leu234 CD2	0.72	2.15	1.37	2.64	3.23	2.83	2.72	0.53	4.44	3.10	3.41
His235 N	1.01	1.92	1.24	3.20	2.91	2.74	3.10	0.54	2.30	0.95	1.37
His235 CA	1.16	2.07	1.27	3.25	2.99	2.98	3.33	0.11	2.50	1.10	1.45
His235 C	0.92	2.11	0.92	3.44	3.16	3.19	3.62	0.39	2.95	1.44	1.73
His235 O	0.63	2.19	0.80	3.31	2.96	2.94	3.62	0.56	3.01	1.23	1.90
His235 CB	1.58	2.30	1.43	2.82	2.58	2.67	3.00	0.44	2.13	0.79	1.08
His235 CG	1.74	2.42	1.38	2.65	2.42	2.46	2.72	0.69	1.58	0.22	0.78
His235 ND1	1.63	2.50	1.34	1.98	1.84	2.46	2.38	2.40	2.65	2.11	1.46
His235 CD2	2.08	2.54	1.35	3.11	2.88	2.18	2.70	1.31	2.02	2.23	0.72
His235 CE1	1.86	2.71	1.29	2.07	2.06	2.19	2.15	2.23	2.15	1.19	0.93
His235 NE2	2.17	2.75	1.30	2.74	2.67	2.02	2.35	1.17	0.73	1.62	0.68
Pro236 N	1.24	2.09	0.91	3.89	3.65	3.62	4.00	0.57	3.33	2.06	1.86
Pro236 CA	1.29	2.09	0.85	4.15	3.88	3.86	4.30	0.86	3.77	2.44	2.17
Pro236 C	1.14	2.05	0.57	3.63	3.45	3.22	3.76	0.35	3.44	1.88	1.81
Pro236 O	1.06	2.23	0.51	3.42	3.32	2.85	3.64	0.66	3.68	1.94	2.06

lrt3	lrt4	lrt5	lrt6	lrt7	lrth	lrti	ltvr	luwb	lvrt	lvru
4.39	6.14	6.03	6.43	6.29	4.20	5.15	6.04	5.84	5.12	5.23
4.59	4.86	4.66	4.91	4.98	4.16	5.09	5.64	5.85	5.03	5.27
4.74	5.09	4.87	4.89	5.28	4.13	4.52	5.16	5.68	4.75	4.86
4.81	4.89	4.61	4.66	4.83	4.15	5.39	5.45	5.99	5.12	5.44
4.96	6.24	6.00	5.82	6.26	4.12	4.81	4.97	5.82	4.85	5.04
5.00	5.81	5.55	5.30	5.73	4.12	5.24	5.11	5.97	5.03	5.33
2.39	2.12	2.06	2.04	2.05	2.08	2.14	3.99	3.77	2.83	2.63
2.33	1.84	1.80	1.78	1.78	1.98	1.94	3.86	3.64	2.72	2.59
2.30	1.67	1.65	1.52	1.45	1.90	1.93	3.98	3.75	2.71	2.57
2.81	2.36	2.38	2.32	2.12	2.54	2.39	4.70	4.53	3.33	3.11
2.88	2.39	2.34	2.29	2.33	2.60	2.17	4.13	3.94	3.13	3.13
2.60	2.13	2.01	2.60	2.59	2.10	2.03	4.09	4.04	2.78	3.24
4.90	4.75	4.66	5.14	4.85	4.59	3.90	5.00	5.50	4.55	3.60
3.00	2.36	2.30	3.11	3.49	2.85	3.68	3.34	3.37	4.33	3.06
1.82	1.05	1.10	0.97	0.69	1.43	1.05	3.33	3.19	2.15	2.03
2.11	1.47	1.58	1.64	1.36	2.01	0.95	3.38	3.43	2.45	2.32
2.71	1.92	2.07	2.11	1.77	2.53	1.14	3.48	3.69	3.06	2.81
2.60	1.57	1.90	1.49	1.14	2.24	1.24	3.35	3.35	3.19	2.75
1.79	1.42	1.56	1.77	1.66	2.05	0.61	3.00	3.10	2.08	1.99
1.36	1.11	1.22	1.30	1.31	1.62	0.48	2.82	2.78	1.45	1.50
3.50	2.20	1.81	3.01	3.38	2.26	1.00	2.12	2.76	2.94	3.39
0.92	1.06	1.04	1.78	1.20	1.23	1.28	3.22	2.51	1.69	1.21
3.22	2.05	1.86	2.16	2.67	2.18	0.54	2.12	2.44	2.31	2.83
0.86	0.56	0.92	1.22	0.86	1.07	1.18	2.79	2.31	0.39	0.33
3.45	2.78	2.80	3.35	3.03	3.38	1.26	3.72	4.34	3.56	3.34
4.06	3.26	3.31	3.89	2.55	3.96	1.51	3.83	4.64	4.20	3.84
3.54	2.77	2.70	3.15	2.73	3.40	1.18	3.46	3.97	3.73	3.46
3.68	2.79	2.64	3.08	2.64	3.50	1.24	3.11	3.66	3.95	3.65

Appendix 7.13 Displacement data for the NNIBP (cont.)

	1rtd	2hmi	1hmv	1bqm	1bqn	1hni	1hnv	1klm	1rev	1rt1	1rt2	1rt3	1rt4	1rt5	1rt6	1rt7	1rth	1rti	1tvr	1uwb	1vrt	1vru
Pro236 CB	1.97	2.15	1.30	4.91	4.45	4.57	4.98	1.38	4.28	3.34	2.64	4.97	4.22	4.30	5.30	5.03	5.11	1.94	4.18	5.39	4.98	4.43
Pro236 CG	2.22	2.66	1.35	4.85	4.68	4.42	4.98	1.27	5.21	3.31	2.39	6.06	5.45	5.44	5.41	5.09	6.30	1.62	4.03	5.05	5.78	5.59
Pro236 CD	1.80	2.11	1.05	4.39	4.31	3.87	4.43	0.76	3.52	2.66	2.02	4.12	3.53	3.46	4.42	4.06	4.25	1.28	3.92	4.63	3.85	3.89
Tyr318 N	0.84	1.26	0.72	1.25	1.15	1.28	1.06	0.23	0.55	0.33	0.26	0.41	0.22	0.27	0.37	0.21	0.23	0.67	1.61	1.34	0.42	0.52
Tyr318 CA	0.59	1.02	0.73	1.39	1.31	1.48	1.31	0.25	0.74	0.46	0.70	0.83	0.37	0.26	0.34	0.37	0.43	0.82	1.66	1.51	0.67	0.71
Tyr318 C	0.55	0.79	0.73	1.30	1.27	1.40	1.19	0.30	0.70	0.41	0.74	0.91	0.32	0.26	0.37	0.44	0.39	0.76	1.51	1.42	0.63	0.68
Tyr318 O	0.54	0.77	0.71	1.30	1.17	1.54	1.05	0.35	0.60	0.38	0.83	0.94	0.22	0.21	0.45	0.48	0.27	0.64	1.40	1.38	0.63	0.66
Tyr318 CB	0.50	1.16	0.79	1.54	1.59	1.68	1.71	0.37	1.02	0.67	0.97	1.08	0.75	0.54	0.59	0.62	0.74	0.98	2.03	1.87	1.08	1.06
Tyr318 CG	0.51	1.30	0.85	1.77	1.87	1.91	2.05	0.52	1.35	0.87	1.03	1.43	0.95	0.78	0.85	0.80	0.95	1.12	2.39	2.22	1.40	1.32
Tyr318 CD1	0.45	1.31	1.83	2.14	2.08	2.33	2.25	0.69	1.53	1.24	1.08	1.63	1.33	1.29	1.16	1.03	1.15	1.21	2.65	2.41	1.62	1.47
Tyr318 CD2	0.63	1.84	2.28	2.08	2.26	3.24	2.36	0.54	1.69	0.93	1.03	1.76	1.20	1.27	0.99	0.88	1.19	1.21	2.94	2.98	1.82	1.73
Tyr318 CE1	0.52	1.50	1.94	2.37	2.49	2.41	2.55	0.85	1.86	1.37	1.10	1.98	1.50	1.47	1.36	1.16	1.35	1.34	2.99	2.72	1.94	1.77
Tyr318 CE2	0.64	1.93	2.22	2.43	2.63	3.59	2.67	0.68	2.01	1.09	1.06	2.08	1.37	1.42	1.22	1.04	1.41	1.34	3.38	3.31	2.11	1.98
Tyr318 CZ	0.57	1.59	0.89	2.40	2.67	2.51	2.67	0.83	2.03	1.18	1.10	2.13	1.34	1.22	1.31	1.11	1.39	1.39	3.22	2.93	2.06	1.92
Tyr318 OH	0.65	1.74	2.91	2.81	3.17	2.88	2.95	0.98	2.38	1.32	1.16	2.49	1.53	1.44	1.51	1.25	1.64	1.52	3.68	3.32	2.38	2.24
Glu138 N	0.64	0.54	0.74	1.05	1.13	1.37	0.88	1.06	1.09	0.86	1.09	2.00	1.07	1.03	0.84	0.87	1.33	1.22	1.17	1.38	0.93	1.51
Glu138 CA	0.65	0.56	0.94	1.27	1.34	1.43	0.95	1.31	1.20	1.18	1.30	2.10	1.25	1.21	1.06	1.24	1.55	1.32	1.34	1.56	1.03	1.68
Glu138 C	1.14	0.66	0.81	1.39	1.30	1.54	0.67	1.17	1.05	1.06	1.11	1.84	1.20	1.11	1.00	1.14	1.51	1.43	1.04	1.24	0.94	2.02
Glu138 O	1.26	0.98	1.00	2.03	1.81	1.80	0.47	1.21	1.00	1.05	1.30	1.57	1.37	1.28	1.16	1.07	1.51	1.48	1.21	1.05	1.04	2.06
Glu138 CB	0.74	0.76	1.41	1.23	1.42	1.39	1.21	1.81	1.62	1.74	1.81	2.29	1.70	1.73	1.58	1.68	1.94	1.39	1.71	1.87	1.47	1.66
Glu138 CG	1.56	0.70	3.81	3.27	3.32	3.36	3.83	4.24	4.23	4.08	4.25	4.61	3.97	4.12	4.17	4.15	4.39	1.83	4.15	4.33	4.06	2.94
Glu138 CD	0.66	1.40	4.14	4.50	3.98	3.19	4.01	4.52	4.64	4.46	4.67	4.66	4.39	4.50	4.54	4.46	4.71	1.61	4.66	4.62	4.45	2.95
Glu138 OE1	1.65	3.31	2.48	4.48	2.45	1.43	3.49	3.13	2.94	2.79	2.87	3.73	2.51	2.60	2.59	6.50	2.79	1.04	3.08	3.28	2.49	1.58
Glu138 OE2	2.30	0.41	6.11	6.25	6.17	5.22	5.27	6.55	6.73	6.65	6.84	5.84	6.65	6.72	6.81	6.70	6.96	3.55	6.77	6.33	6.74	5.05

The measured distances between the unliganded (1dlo) structure and each RT structure for all the atoms of the residues of NNIBP.

Appendix 7.14 Displacement data between motifs of RT

	$73C_\alpha$:$231C_\alpha$	$91C_\alpha$:$185C_\alpha$	$184C_\alpha$:$498C_\alpha$	$498C_\alpha$:$231C_\alpha$
1dlo	25.23	16.32	60.93	58.37
1hmv		16.27	62.12	60.47
1rtd	23.22	12.57	61.01	58.01
2hmi	27.61	14.16	61.11	59.49
1bqm	28.41	13.14	63.04	55.91
1bqn	28.03	12.66	62.39	56.51
1hni	28.28	12.97	62.41	55.12
1hnv	29.33	12.77	62.86	55.28
1klm	30.54		62.37	56.26
1rev	31.79	16.05	62.45	55.11
1rt1	31.65	12.86	62.73	54.81
1rt2	30.43	13.89	62.43	54.98
1rt3	31.28	16.96	61.32	54.95
1rt4	31.69	12.80	63.38	55.74
1rt5	32.11	12.86	63.24	55.42
1rt6	31.47	13.26	62.36	54.96
1rt7	31.59	13.19	62.73	55.20
1rth	30.13	12.95	63.59	56.24
1rti	29.39	13.61	63.04	56.80
1tvr	29.10	13.15	63.43	55.86
1uwb	28.93	12.86	62.83	55.43
1vrt	31.15	11.98	62.48	55.26
1vru	31.05	13.01	63.07	55.41

The measured distance between selected motifs in each of the crystal structures.

73 C_α of the fingers segment of the template grip and 231 C_α of the primer grip
91 C_α of the palm segment of the template grip and 185 C_α of the PAS
184 C_α of the PAS and 498 C_α of the RNase H active site
498 C_α of the RNase H active site and 231 C_α of the primer grip

Appendix 7.15 Hypoedit utility in Catalyst

The Hypoedit file of Catalyst used to manually construct a pharmacophore named 'crystalQuery.chm'.

```
!/bin/csh
#
# Here is an example of how one would use the hypoedit program to make
a
# Catalyst 3D query named crystalQuery.chm that could be used to
search a
# Catalyst database.  Note that all coordinates and tolerances are in
# picometers.
#
$CATALYST_GOODIES/hypoedit \
      -dict $CATALYST_CONF/Dictionary.chm \
      -feat HYDROPHOBIC \
      -feat HYDROPHOBIC \
      -feat HYDROPHOBIC \
      -feat "HB DONOR" \
      -feat "HB ACCEPTOR" \
      -blob  100.0  100.0  100.0 150.0 1 \
      -blob  200.0  200.0  200.0 150.0 2 \
      -blob  300.0  300.0  300.0 150.0 3 \
      -blob  400.0  400.0  400.0 150.0 4 \
      -blob  500.0  500.0  500.0 200.0 5 \
      -blob  600.0  600.0  600.0 150.0 6 \
      -blob  700.0  700.0  700.0 200.0 7 \
      -xvol  800.0  800.0  800.0 250.0 \
      -xvol  900.0  900.0  900.0 250.0 \
      -xvol 1000.0 1000.0 1000.0 250.0 \
      -xvol 1100.0 1100.0 1100.0 250.0 \
      crystalQuery.chm
```

Appendix 7.15.1 Hypoedit file of activity 'weighted' structural query

Hypoedit file used to generate a Catalyst pharmacophore, named 'weighted_pocket.chm' of the Unity 3D structural query based on the 'weight' binding pocket.

```
-dict $CATALYST_CONF/Dictionary.chm \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB ACCEPTOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat "HB DONOR" \
-feat HYDROPHOBIC \
-feat HYDROPHOBIC \
-feat HYDROPHOBIC \
-feat HYDROPHOBIC \
-feat HYDROPHOBIC \
-feat HYDROPHOBIC \
-feat HYDROPHOBIC \
-feat HYDROPHOBIC \
-feat HYDROPHOBIC \
-feat HYDROPHOBIC \
```

```
-blob  -552.87 -51.57 494.66 150.0 1 \
-blob  -530.85 113.81 743.99 250.0 2 \
-blob  930.99 256.05 -136.16 150.0 3 \
-blob  858.95 -35.09 -143.21 250.0 4 \
-blob  502.25 -728.66 -294.81 150.0 5 \
-blob  741.95 -632.79 -162.71 250.0 6 \
-blob  -822.18 -327.38 369.09 150.0 7 \
-blob  -918.05 -47.89 317.19 250.0 8 \
-blob  -431.50 -182.48 -31.69 150.0 9 \
-blob  -480.65 -25.89 -282.81 250.0 10 \
-blob  -342.70 197.06 288.47 150.0 11 \
-blob  -176.95 307.11 77.49 250.0 12 \
-blob  -255.80 148.30 -1190.24 150.0 13 \
-blob  -222.25 -56.19 -973.31 250.0 14 \
-blob  -404.50 62.41 -457.82 150.0 15 \
-blob  -394.15 126.51 -750.71 250.0 16 \
-blob  -261.76 -44.49 -751.04 150.0 17 \
-blob  -276.25 -200.99 -495.51 250.0 18 \
-blob  -285.44 691.11 -73.89 150.0 19 \
-blob  -12.15 748.61 -152.01 250.0 20 \
-blob  -9.48 -353.59 843.95 150.0 21 \
-blob  58.15 -636.69 771.29 250.0 22 \
-blob  -219.95 -409.75 693.08 150.0 23 \
-blob  -5.35 -499.49 519.89 250.0 24 \
-blob  -53.69 -910.81 328.91 150.0 25 \
-blob  -125.05 -649.29 225.89 250.0 26 \
-blob  -463.01 -759.44 415.69 150.0 27 \
-blob  -474.55 -558.99 638.59 250.0 28 \
-blob  -175.62 -700.51 -409.40 150.0 29 \
-blob  -168.55 -500.19 -632.61 250.0 30 \
-blob  -778.61 91.53 562.49 150.0 31 \
-blob  -682.45 255.91 794.29 250.0 32 \
-blob  581.62 -295.64 -477.80 150.0 33 \
-blob  874.25 -247.99 -432.01 250.0 34 \
-blob  -626.69 375.24 501.68 150.0 35 \
-blob  -471.75 425.11 741.69 250.0 36 \
-blob  -807.41 -11.52 698.09 150.0 37 \
-blob  -604.35 181.51 772.99 250.0 38 \
-blob  138.65 524.86 135.02 150.0 39 \
-blob  415.05 461.71 195.99 250.0 40 \
-blob  634.05 -173.63 -491.93 150.0 41 \
-blob  845.95 -203.09 -687.71 250.0 42 \
-blob  -1030.31 -177.08 -147.21 150.0 43 \
-blob  -1199.95 -279.39 64.59 250.0 44 \
-blob  146.87 -433.90 -1008.27 150.0 45 \
-blob  -66.95 -419.09 -812.91 250.0 46 \
-blob  -254.33 -688.46 -410.56 150.0 47 \
-blob  -269.25 -399.29 -394.41 250.0 48 \
-blob  -458.06 -98.31 15.55 150.0 49 \
-blob  -522.05 167.41 -81.41 250.0 50 \
-blob  -159.40 -382.06 -828.64 150.0 51 \
-blob  -217.55 -106.19 -760.71 250.0 52 \
-blob  -638.52 223.04 -497.78 150.0 53 \
-blob  -411.25 43.71 -480.81 250.0 54 \
-blob  -843.18 -293.65 -163.37 150.0 55 \
-blob  -979.45 -38.39 -144.11 250.0 56 \
-blob  -804.77 242.98 122.12 150.0 57 \
-blob  -871.15 20.81 296.29 250.0 58 \
-blob  531.74 923.86 -399.01 150.0 59 \
-blob  368.85 1081.71 -218.31 250.0 60 \
-blob  -393.76 -402.59 24.39 150.0 61 \
```

```
-blob  -123.95 -297.89 42.89 250.0 62 \
-blob  149.51 -548.03 -14.28 150.0 63 \
-blob  411.25 -423.59 -24.61 250.0 64 \
-blob  -12.53 906.82 -797.30 150.0 65 \
-blob  -178.65 732.21 -636.01 250.0 66 \
-blob  -121.49 -96.34 56.92 150.0 67 \
-blob  -28.45 -81.49 331.19 250.0 68 \
-blob  -740.21 -455.84 644.86 150.0 69 \
-blob  -674.05 -730.29 711.19 250.0 70 \
-blob  -495.77 -268.24 843.93 150.0 71 \
-blob  -481.35 -493.69 662.09 250.0 72 \
-blob  -377.26 -342.08 721.87 150.0 73 \
-blob  -241.65 -403.09 472.89 250.0 74 \
-blob  -2.35 -349.42 791.82 150.0 75 \
-blob  240.65 -451.89 671.19 250.0 76 \
-blob  365.82 -111.68 843.39 150.0 77 \
-blob  109.25 0.31 767.69 250.0 78 \
-blob  723.47 -478.30 -119.87 150.0 79 \
-blob  524.85 -302.99 -1.91 250.0 80 \
-blob  -847.97 -215.56 1397.14 150.0 81 \
-blob  -908.15 -247.45 1194.43 150.0 82 \
-blob  -159.35 135.18 -129.94 150 83 \
-blob  166.77 693.64 -86.71 150 84 \
-blob  72.89 779.65 -197.75 150 85 \
-blob  -405.78 713.76 -21.28 150 86 \
-blob  1041.68 216.08 -627.94 150 87 \
-blob  146.52 645.41 -1095.73 150 88 \
-blob  -167.63 461.01 -678.73 150 89 \
-blob  758.60 986.79 -3.54 150 90 \
weighted_pocket.chm
```